

On the Relationship between Mutual Information and Minimum Mean-Square Errors in Stochastic Dynamical Systems

Francisco J. Piera*

Patricio Parada†

October 4, 2007

Abstract

We consider a general stochastic input-output dynamical system with output evolving in time as the solution to a functional coefficients, Itô's stochastic differential equation, excited by an input process. This general class of stochastic systems encompasses not only the classical communication channel models, but also a wide variety of engineering systems appearing through a whole range of applications. For this general setting we find analogous of known relationships linking input-output mutual information and minimum mean causal and non-causal square errors, previously established in the context of additive Gaussian noise communication channels. Relationships are not only established in terms of time-averaged quantities, but also their time-instantaneous, dynamical counterparts are presented. The problem of appropriately introducing in this general framework a signal-to-noise ratio notion expressed through a signal-to-noise ratio parameter is also taken into account, identifying conditions for a proper and meaningful interpretation.

Index Terms. Stochastic dynamical systems, stochastic differential equations (SDE), mutual information, minimum mean square errors (MMSE), non-linear estimation, smoothing, optimal filtering.

*F.J. Piera is with the Department of Electrical Engineering, University of Chile, Av. Tupper 2007, Santiago, 8370451, Chile (e-mail: fpiera@ing.uchile.cl).

†P. Parada is with the Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, 1406 W. Green St., Urbana, IL, 61801-2918 USA, and the Department of Electrical Engineering, University of Chile, Av. Tupper 2007, Santiago, 8370451, Chile (e-mail: paradasa@uiuc.edu).

1 Introduction

Consider the widely used communication system model known as the standard additive white Gaussian noise channel, described by

$$Y_t^r = \sqrt{r} \int_0^t X_s ds + W_t, \quad t \in [0, T], \quad (1)$$

where $r \in [0, \infty)$ is the signal-to-noise ratio parameter, $T \in (0, \infty)$ is a fixed time-horizon, $X = (X_t)_{t \in [0, T]}$ is the transmitted random signal or channel input, $W = (W_t)_{t \in [0, T]}$ is an independent standard Brownian motion or Wiener process representing the noisy transmission environment, and $Y^r = (Y_t^r)_{t \in [0, T]}$ is the received random signal or channel output, corresponding to the respective value of the signal-to-noise ratio parameter r .

Of central importance from an information theoretical point of view is the input-output mutual information, i.e., the mutual information between the processes X and Y^r , denoted by $I(r)$. (Precise mathematical definitions are deferred to the next section.) On the other hand, of central importance from an estimation theoretical point of view are the causal and non-causal minimum mean square errors, in estimating or smoothing X at time $t \in [0, T]$, denoted by $\text{cmmse}_X(t, r)$ and $\text{ncmmse}_X(t, r)$, respectively. Input-output mutual information encloses a measure of how much coded information can be reliably transmitted through the channel for the given input source, whereas the causal and non-causal minimum mean square errors indicate the level of accuracy that can be reached in the estimation of the transmitted message at the receiver, based on the causal or noncausal observation of an output sample path, respectively.

Interesting results on the relationship between filter maps and likelihood ratios in the context of the additive white Gaussian noise channel have been available in the literature for a while (see for example [1] and references therein). An interesting specific result linking information theory and estimation theory in this same Gaussian channel context, concretely, input-output mutual information and causal minimum mean square error, is Duncan's theorem [2] stating, under appropriate finite average power conditions, the relationship

$$I(r) = \frac{r}{2} \int_0^T \text{cmmse}_X(s, r) ds, \quad r \in [0, \infty), \quad (2)$$

i.e., after dividing both sides by T , stating the proportionality (through the factor $\frac{r}{2}$) of mutual information rate per unit time and time average causal minimum mean square error. It was recently shown by Guo et al. [3] that the previous relationship is not the only linking property between information theory and estimation theory in this Gaussian chan-

nel setting, but also that there exists an important result involving input-output mutual information and non-causal minimum mean square error, namely

$$\frac{d}{dr}I(r) = \frac{1}{2} \int_0^T \text{ncmmse}_X(s, r) ds, \quad r \in [0, \infty). \quad (3)$$

As pointed out by Guo et al. [3], an interesting relationship between causal and non-causal minimum mean square errors can then be directly deduced from (2) and (3), giving

$$\int_0^T \text{cmmse}_X(s, r) ds = \frac{1}{r} \int_0^r \int_0^T \text{ncmmse}_X(s, u) ds du, \quad r \in (0, \infty), \quad (4)$$

i.e., after dividing as before both sides by T , stating the equality between time average causal minimum mean square error and the in turn averaged over the signal-to-noise ratio, time average non-causal minimum mean square error. Equations (2) to (4) can for example be used to study asymptotics of input-output mutual information and minimum mean square errors, and to find new representations of information measures [3].

An increasing necessity of considering general stochastic models has arisen during the last decades in the stochastic systems modelling community, not just from a communication systems standpoint, but from a wide variety of applications demanding the consideration of general stochastic input-output dynamical systems described by Itô's stochastic differential equations of the form

$$Y_t^r = \sqrt{r} \int_0^t F(s, X, Y^r) ds + \int_0^t G(s, Y^r) dW_s, \quad t \in [0, T], \quad (5)$$

with X the input stochastic process to the system, r a non-negative real parameter (to be interpreted further in subsequent sections), Y^r the corresponding system output stochastic process¹, and F and G given (time-varying) non-anticipative functionals, i.e., with $F(t, X, Y^r)$ depending on the random paths of X and Y^r only up to time t , and similarly for $G(t, Y^r)$. Note since W is an infinite variation process, the integral

$$\int_0^t G(s, Y^r) dW_s$$

is an Itô's stochastic integral and not an standard pathwise Lebesgue-Stieltjes integral. For the input process X , the corresponding system output Y^r evolves in time then as the solution to the stochastic differential equation (5). (Once again, we defer mathematical preciseness to subsequent sections.) From a modelling point of view, the flexibility offered

¹To ease notation we simply write Y^r , instead of for example $Y^{r,X}$, the input process X being clear from the context.

by the general model (5) captures a vast collection of system output stochastic behaviors, as for example the class of strong Markov processes [4]. As mentioned, general stochastic input-output dynamical systems as the one portrayed by (5) appear in a wide variety of stochastic modelling applications. They are usually obtained by a weak-limit approximation procedure, where a sequence of properly scaled and normalized subjacent stochastic models is considered and shown to converge, in a weak or in distribution stochastic process convergence sense [5–8], to the solution of a corresponding stochastic differential equation. Just to name a few, some examples are applications to adaptive antennas, channel equalizers, adaptive quantizers, hard limiters, and synchronization systems such as standard phase-locked loops and phase-locked loops with limiters [8]. They have also become extremely useful in heavy-traffic approximations of stochastic networks of queues in operations research and communications [6, 9–15], where they are usually brought into the picture along with the Skorokhod (or reflection) map constraining a given process to stay inside a certain domain or spatial region [6, 16], and in mathematical economics (option pricing and the Black-Scholes formula, arbitrage theory, consumption and investment problems, insurance and risk theory, etc.) and stochastic control theory [17–20].

The so obtained diffusion² models offer two main modelling advantages. On one hand, they usually wash off in the limit non fundamental model details, accounting for mathematical tractability and leading to a diffusion model that captures the main aspects and trade offs involved. On the other, they have the enormous advantage of taking the modelling setting to the stochastic analysis framework, where the whole machinery of stochastic calculus is available.

From a purely communication systems modelling viewpoint, it is worth emphasizing that a general stochastic input-output dynamical system such as (5) encompasses all standard communication Gaussian channel models as particular cases, such as the white Gaussian noise channel (with/without feedback) or its extension to the colored Gaussian noise case. These particular instances will be mathematically described in subsequent sections. It is also worth mentioning that though more sophisticated mathematical frameworks have been considered in the literature, as for example an infinite dimensional Gaussian setting [21] with the associated Malliavin’s stochastic analysis tools [22, 23], the essentially white Gaussian nature of the noise has remained untouched by most. In this regard, the main tools considered to establish relationships such as (3) and (4) usually depend critically on a Lévy structure³ for the noisy term⁴ and, specifically, on its independent increment property

²An strong Markov process with continuous sample paths is generally termed a *diffusion*.

³Recall a process with stationary independent increments is termed a Lévy process.

⁴Following the communication systems jargon, we refer to the integral $\int_0^t G(s, Y^r) dW_s$ as the noise term. Further interpretations on this line are discussed in the next section.

such as in the purely Brownian motion noisy term case where⁵ $G \equiv \alpha \in \mathbb{R}$ (a constant) in (5). The flexibility of an Itô's stochastic integral with general functional G in (5) allows for a much generality of stochastic behaviors, including non-Lévy ones.

The main objective of this paper is to establish links between information theory and estimation theory in the general setting of a stochastic input-output dynamical system described by (5). Specifically, it is shown that an analogous relationship to (2) can be written in this setting, so extending classical Duncan's theorem for standard additive white Gaussian noise channels with and without feedback [2, 24] to this generalized model. Proofs are in the framework of absolutely continuity properties of stochastic process measures, subjacent to the Girsanov's theorem [4, 25]. Relationships (3) and (4) are also studied in this generalized setting. As mentioned, they were shown to hold in the context of the additive white Gaussian noise channel in the work of Guo et al. [3]. However, as also pointed out in that work, they fail to hold when feedback is allowed in that purely Gaussian noise framework. We show that failure obeys to the fact that a proper notion of a signal-to-noise ratio expressed through a parameter such as r in (1) cannot be properly introduced in that case, and, by adequately identifying conditions for a signal-to-noise ratio parameter to have a meaningful interpretation, we find analogous relationships to (3) and (4) holding for a subclass of models contained in the general setting of (5). The analysis includes the identification and proper definition of three important classes of related systems, namely what we will come to call *quasi-signal-to-noise*, *signal-to-noise* and *strong-signal-to-noise systems*.

Another particular aspect adding scope of applicability to the results exposed in the present paper, in addition to the system model generality considered here, is related to the fact that not only relationships involving time-averaged quantities such as in (2) and (3) above are extended to this general setting, but also time-instantaneous counterparts are provided. This fact brings dynamical relationships into the picture, allowing to write general integro-partial-differential equations characterizing the different information and estimation theoretic quantities involved. Dynamical relationships are usually absent in the information theory context, being in general difficult to find. The results provided extend then not only the traditional Gaussian system framework, but also the customary time-independent, static relationships setting where information and estimation theoretic quantities are studied for stationary (usually Gaussian) system input processes [26–28], or for non-stationary system inputs but in terms of time-averaged quantities [2, 24].

Finally, we mention that for sake of simplicity in the exposition of the results we will

⁵The process $(\int_0^t G(s, Y^r) dW_s)_{t \in [0, T]}$ is not a Lévy process unless G is a fixed constant.

consider throughout the paper one-dimensional systems and processes. However, all the results presented in the paper have indeed multi-dimensional counterparts. These and further possible extensions, with the corresponding related generalized results, will not be difficult to carry out by the reader in light of the computations developed in the paper, and therefore we will only mention the main ideas involved by the end of the paper without giving corresponding proofs.

The organization of the paper is as follows. In Section 2 we introduce the mathematically rigorous system model setup, including the model definition, the main general assumptions, and the different information and estimation theoretic quantities involved, such as input-output mutual information and causal and non-causal minimum mean square errors, as well as important concepts from the general theory of stochastic process such as the absolutely continuity of stochastic process measures. In Section 3 we establish the relationship linking input-output mutual information and causal minimum mean square error for the general dynamical input-output stochastic system considered in the paper, generalizing the known result for the standard additive white Gaussian noise channel with/without feedback. In Section 4 we identified conditions under which a proper notion of a signal-to-noise ratio parameter can be introduced in our general system setting. We distinguish three major subclasses of systems and give appropriate characterizations. In Section 5 we establish the corresponding generalization of the relationship linking input-output mutual information and non-causal minimum mean square error for an appropriate subclass of system models. In Section 6 we provide the corresponding time-instantaneous counterparts of the previous results. In Section 7 we comment on further model extensions and related results. Finally, in Section 8 we briefly comment on the scope of the results exposed.

2 Preliminary Elements

This section provides the precise mathematical framework upon which the present work is elaborated. In addition to introduce a thoroughly mathematical definition of the dynamical system model to be considered throughout, it also introduces the main concepts from information theory and statistical signal processing appearing in subsequent sections, such as the notion of mutual information between stochastic processes, the accompanying notion of absolutely continuity of measures induced by stochastic processes, and minimum-mean square errors in estimating and smoothing stochastic processes.

2.1 System Model Definition

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, $T \in (0, \infty)$ be fixed throughout, and $(\mathcal{F}_t)_{t \in [0, T]}$ be a filtration on \mathcal{F} , i.e., a nondecreasing family of sub- σ -algebras of \mathcal{F} . We assume the filtration $(\mathcal{F}_t)_{t \in [0, T]}$ satisfies the usual hypotheses [4], i.e., \mathcal{F}_0 contains all the \mathbb{P} -null sets of \mathcal{F} and $(\mathcal{F}_t)_{t \in [0, T]}$ is right-continuous. Also, let $W = (W_t, \mathcal{F}_t)_{t \in [0, T]}$ be a one-dimensional standard Brownian motion⁶ [17], and (C_T, \mathcal{B}_T) be the measurable space of functions in C_T , the space of all functions $f : [0, T] \rightarrow \mathbb{R}$ continuous on $[0, T]$, equipped with the σ -algebra \mathcal{B}_T of finite-dimensional cylinder sets in C_T [17], i.e.⁷,

$$\mathcal{B}_T \doteq \sigma \left(\left\{ C_{\{t_i\}_{i=1}^n}^\Gamma : n \in \mathbb{Z}_+, \{t_i\}_{i=1}^n \subseteq [0, T], \Gamma \in \mathcal{B}(\mathbb{R}^n) \right\} \right)$$

where $\mathcal{B}(\mathbb{R}^n)$ denotes the collection of Borel sets in \mathbb{R}^n , $n \in \mathbb{Z}_+ \doteq \{1, 2, \dots\}$, and

$$C_{\{t_i\}_{i=1}^n}^\Gamma \doteq \{f \in C_T : (f(t_1), \dots, f(t_n)) \in \Gamma\}$$

for each $n \in \mathbb{Z}_+$, $\{t_i\}_{i=1}^n \subseteq [0, T]$, and $\Gamma \in \mathcal{B}(\mathbb{R}^n)$. In a similar way we introduce, for each $t \in [0, T]$, the σ -algebra \mathcal{B}_t of finite-dimensional cylinder sets in the space C_t of all functions $f : [0, t] \rightarrow \mathbb{R}$ continuous on $[0, t]$, and, for A_T a given family of functions $f : [0, T] \rightarrow \mathbb{R}$, the σ -algebras \mathcal{B}_{A_T} and \mathcal{B}_{A_t} of finite-dimensional cylinder sets in A_T and A_t , respectively, with

$$A_t \doteq \left\{ f|_{[0, t]} : f \in A_T \right\}$$

and $f|_{[0, t]}$ the restriction of $f : [0, T] \rightarrow \mathbb{R}$ to the subinterval $[0, t]$.

For each $r \in \mathbb{R}_+ \doteq [0, \infty)$ we consider a stochastic process $Y^r = (Y_t^r, \mathcal{F}_t)_{t \in [0, T]}$, with paths or trajectories in the measurable space (C_T, \mathcal{B}_T) , and having Itô's stochastic differential

$$dY_t^r = \sqrt{r}F(t, X, Y^r)dt + G(t, Y^r)dW_t \quad (6)$$

with $Y_0^r = 0$, where

- the stochastic process $X = (X_t, \mathcal{F}_t)_{t \in [0, T]}$, with trajectories in a given measurable space of functions (A_T, \mathcal{B}_{A_T}) , is independent of W , and
- the functionals $F : [0, T] \times A_T \times C_T \rightarrow \mathbb{R}$ and $G : [0, T] \times C_T \rightarrow \mathbb{R}$ are measurable

⁶The notation $(Z_t, \mathcal{F}_t)_{t \in [0, T]}$ indicates the stochastic process $(Z_t)_{t \in [0, T]}$ is $(\mathcal{F}_t)_{t \in [0, T]}$ -adapted, i.e., Z_t is \mathcal{F}_t -measurable for each $t \in [0, T]$. In case of a Brownian motion $W = (W_t, \mathcal{F}_t)_{t \in [0, T]}$, it also indicates W is a martingale on that filtration, coinciding then with the also called in the literature Wiener process relative to $(\mathcal{F}_t)_{t \in [0, T]}$ [25].

⁷We write, as usual, $\sigma(\cdot)$ for the corresponding generated σ -algebra.

and non-anticipative, i.e., they are $\sigma(\mathcal{B}([0, T]) \times \mathcal{B}_{A_T} \times \mathcal{B}_T)$ - and $\sigma(\mathcal{B}([0, T]) \times \mathcal{B}_T)$ -measurable⁸, respectively, and, for each $t \in [0, T]$, $F(t, \cdot, \cdot)$ and $G(t, \cdot)$ are $\sigma(\mathcal{B}_t \times \mathcal{B}_{A_t})$ - and \mathcal{B}_t -measurable, respectively as well. In other words, the functionals F and G are jointly measurable with respect to (w.r.t.) all their corresponding arguments, and depend at each time $t \in [0, T]$ on $f \in A_T$ and $g \in C_T$ only through $f|_{[0, t]}$ and $g|_{[0, t]}$, i.e., only on the pieces of trajectories

$$\{f(s), g(s) : s \in [0, t]\}.$$

Conditions for properly interpreting $r \in \mathbb{R}_+$ as a signal-to-noise ratio (SNR) parameter for system (6) will be discussed in Section 4.

As discussed in Section 1, we may interpret equation (6) as a general stochastic input-output dynamical system with input stochastic process X and output stochastic process Y^r , for each given value of the parameter r , the output process Y^r evolving in time $t \in (0, T]$ as an Itô's process [29] with differential given by (6). Though the scope of applicability of a general dynamical system model such as (6) exceeds by far a purely communication system setting, it is worth mentioning that from a classical communication channels point of view we shall interpret X as a random input message being printed in the “channel signal component” $\sqrt{r}Fdt$, received at the channel output embedded in the additive “channel noisy term” GdW_t . The standard additive white Gaussian noise channel (AWGNC) being obtained from (6) by taking

$$F(t, f, g) = f(t) \text{ and } G(t, g) \equiv 1,$$

for each $t \in [0, T]$, $f \in A_T$, and $g \in C_T$, i.e., with the corresponding output process or “random received signal” Y^r evolving for $t \in (0, T]$ according to

$$dY_t^r = \sqrt{r}X_t dt + dW_t, \tag{7}$$

and $r \in \mathbb{R}_+$ the channel SNR⁹. In this same line, note when G in (6) is allowed to depend only on $t \in [0, T]$, and not on Y^r , the noisy term

$$\int_0^t G(s) dW_s, \quad t \in [0, T],$$

⁸Similarly than for \mathbb{R}^n , $\mathcal{B}([0, T])$ denotes the collection of Borel sets in the interval $[0, T]$.

⁹The interpretation of r as an SNR parameter is discussed at full in Section 4.

is a zero-mean Gaussian process with covariance function given by [30]

$$\mathbb{E} \left[\int_0^{t_1} G(s) dW_s \int_0^{t_2} G(s) dW_s \right] = \int_0^{\min\{t_1, t_2\}} G^2(s) ds,$$

$t_1, t_2 \in [0, T]$, provided G is square-integrable on $[0, T]$, i.e.,

$$\int_0^T G^2(s) ds < \infty.$$

This case is usually known in the literature as the additive colored Gaussian noise channel.

It is technically suitable to treat W in (6) as a system input too, as it is sometimes the case when the stochastic system at hand is obtained by a weak limit procedure of a properly scaled and normalized sequence of subjacent system models [8, 13]. The *principle of causality* for dynamical systems [17] requires the output process Y^r at time $t \in [0, T]$, Y_t^r ($Y_0^r = 0$), to depend only on the values

$$\{X_s, W_s : s \in [0, t]\},$$

i.e., only on the past history of X and W up to time t . (This requirement finds a precise mathematical expression in the adaptability condition (I) imposed below.) Therefore the non-anticipability nature imposed on the functional F and G .

For a fixed deterministic trajectory $x(\cdot) \in A_T$ in place of X in (6), we have the corresponding output stochastic process, denoted as $Y^{r,x}$ for each r , evolving as a solution of the stochastic differential equation (SDE) [4]

$$Y_t^{r,x} = \sqrt{r} \int_0^t F(s, x, Y^{r,x}) ds + \int_0^t G(s, Y^{r,x}) dW_s, \quad (8)$$

$t \in [0, T]$. When for each $t \in [0, T]$ and $g \in C_T$ we have $F(t, x, g) = \overline{F}_x(t, g(t))$ and $G(t, g) = \overline{G}(t, g(t))$, for some Borel-measurable functions $\overline{F}_x : [0, T] \times \mathbb{R} \rightarrow \mathbb{R}$ and $\overline{G} : [0, T] \times \mathbb{R} \rightarrow \mathbb{R}$, $Y^{r,x}$ is indeed a diffusion process, i.e., an strong Markov process with continuous sample paths on $[0, T]$ [31]. Though we are of course interested in the general case when the input to the system is a stochastic process X as in (6), rather than a fix trajectory x as in (8), we refer to (6) as an SDE system motivated from the above discussion. In fact, for X and Y^r related as in (6), we may look at Y^r as solving the SDE with random drift coefficient

$$Y_t^r = \sqrt{r} \int_0^t B_X(\omega, s, Y^r) ds + \int_0^t G(s, Y^r) dW_s,$$

$t \in [0, T]$, where the random drift functional $B_X : \Omega \times [0, T] \times C_T \rightarrow \mathbb{R}$ is given by

$$B_X(\omega, t, g) \doteq F(t, X(\omega), g) \quad (9)$$

for each $t \in [0, T]$ and $g \in C_T$. Note that B_X is not only $\sigma(\mathcal{F} \times \mathcal{B}([0, T]) \times \mathcal{B}_T)$ -measurable, but also, for each $t \in [0, T]$, $B_X(\cdot, t, \cdot)$ is $\sigma(\mathcal{F}_t^X \times \mathcal{B}_t)$ -measurable, where

$$\mathcal{F}_t^X \doteq \sigma(\{X_s : s \in [0, t]\}),$$

$t \in [0, T]$, is the history of X up to time t , i.e., the minimal σ -algebra on Ω making all the random variables $\{X_s : s \in [0, t]\}$ measurable.

Throughout we shall assume the following conditions are satisfied.

- (I) For each $r \in \mathbb{R}_+$ the stochastic process Y^r is the pathwise unique strong solution of equation (6) [32, 33]. It is strong in the sense that, for each $t \in [0, T]$, Y_t^r is measurable w.r.t. the σ -algebra

$$\mathcal{F}_t^{X,W} \doteq \sigma(\{X_s, W_s : s \in [0, t]\}),$$

which represents the joint history of X and W up to time t , i.e., the minimal σ -algebra on Ω making all the random variables $\{X_s, W_s : s \in [0, t]\}$ measurable. Equivalently, the stochastic process Y^r is adapted to the filtration $(\mathcal{F}_t^{X,W})_{t \in [0, T]}$. It is pathwise unique in the sense that if Y^r and \tilde{Y}^r are two strong solutions of (6), then $Y_t^r = \tilde{Y}_t^r$ for all $t \in [0, T]$, \mathbb{P} -almost surely, i.e.,

$$\mathbb{P}\left(Y_t^r = \tilde{Y}_t^r, t \in [0, T]\right) = 1.$$

(See Remark 2.1 below for the existence and uniqueness of such a solution.)

- (II) The non-anticipative functionals F and G are such that

$$\int_0^T |F(t, f, g)| dt < \infty \text{ and } \int_0^T G^2(t, g) dt < \infty,$$

for each $f \in A_T$ and $g \in C_T$.

- (III) For each $t \in [0, T]$ and $f, g \in C_T$,

$$|G(t, f) - G(t, g)|^2 \leq K_1 \int_0^t |f(s) - g(s)|^2 dL(s) + K_2 |f(t) - g(t)|^2, \quad (10)$$

$$G^2(t, f) \leq K_1 \int_0^t (1 + f^2(s)) dL(s) + K_2 (1 + f^2(t)), \quad (11)$$

and

$$G^2(t, f) \geq K > 0, \quad (12)$$

where $L : [0, T] \rightarrow \mathbb{R}$ is a non-decreasing, right-continuous function satisfying $L(t) \in [0, 1]$ for each $t \in [0, T]$, and K , K_1 and K_2 are finite constants. Equations (10), (11) and (12) correspond to Lipschitz, linear growth and non-degeneracy conditions on the non-anticipative functional G , respectively.

(IV) For each $r \in \mathbb{R}_+$,

$$\mathbb{P} \left(\int_0^T F^2(t, X, Y^r) dt < \infty \right) = \mathbb{P} \left(\int_0^T F^2(t, X, \xi) dt < \infty \right) = 1,$$

where $\xi = (\xi_t, \mathcal{F}_t)_{t \in [0, T]}$ is the pathwise unique strong solution of the equation

$$d\xi_t = G(t, \xi_t) dW_t, \quad \xi_0 = 0.$$

(Existence and uniqueness of ξ follow from condition (III) and [25, Theorem 4.6, p.128].)

(V) For each $r \in \mathbb{R}_+$,

$$\int_0^T \mathbb{E} [|F(t, X, Y^r)|] dt < \infty \quad (13)$$

and

$$\mathbb{P} \left(\int_0^T \mathbb{E}^2 [F(t, X, Y^r) | \mathcal{F}_t^{Y^r}] dt < \infty \right) = 1,$$

where, for each $r \in \mathbb{R}_+$ and $t \in [0, T]$,

$$\mathcal{F}_t^{Y^r} \doteq \sigma(\{Y_s^r : s \in [0, t]\}),$$

the history of Y^r up to time t . Here, and throughout, $\mathbb{E}[\cdot | \cdot]$ denotes conditional expectation, as usual.

Remark 2.1. *If the random drift functional B_X in (9) satisfies appropriate similar Lipschitz and linear growth conditions as to G in (III), in a \mathbb{P} -almost surely basis of course and with K_1 and K_2 and L random variables and stochastic process, respectively, then the existence of a pathwise unique strong solution of (6) can be read off from [25, Theorem 4.6, p.128]. We do not explicitly require such conditions though, but just assume the corresponding existence and uniqueness in condition (I).*

Remark 2.2. *As the reader will easily verify, all the results in the paper hold if condition (I) is weakened to just ask that, for each $r \in \mathbb{R}_+$, Y^r is any strong solution of (6), i.e., to just assume the existence of each Y^r as any given strong solution of equation (6). We demand uniqueness in condition (I) for sake of preciseness, as well as to properly interpret (6) as an input-(unique)output dynamical system.*

As it will be detailed in subsequent sections, conditions (I) to (V), as well as the assumption on the stochastic independence of processes X and W , ensure the existence of several densities or Radon-Nikodym derivatives between the measures induced by the stochastic processes involved in their corresponding sample spaces of functions. These Radon-Nikodym derivatives are introduced in the following subsection.

2.2 Absolutely Continuity of Stochastic Process Measures

Recall from the previous subsection that the stochastic processes $X = (X_t)_{t \in [0, T]}$ and $Y^r = (Y_t^r)_{t \in [0, T]}$ (each $r \in \mathbb{R}_+$) have trajectories, or sample paths, in the measurable spaces of functions (A_T, \mathcal{B}_{A_T}) and (C_T, \mathcal{B}_T) , respectively. In the same way, the auxiliary process $\xi = (\xi_t)_{t \in [0, T]}$, introduced previously in condition (IV), has sample paths in the measurable space (C_T, \mathcal{B}_T) . We denote by

$$\mu_X, \mu_{Y^r} \text{ and } \mu_\xi$$

the corresponding measures they induced in the measurable spaces (A_T, \mathcal{B}_{A_T}) , (C_T, \mathcal{B}_T) , and (C_T, \mathcal{B}_T) , respectively. Analogously, we denote by

$$\mu_{X, Y^r}$$

the (joint) measure induced by the pair of processes (X, Y^r) in the measurable space $(A_T \times C_T, \sigma(\mathcal{B}_{A_T} \times \mathcal{B}_T))$.

As it was mentioned by the end of the previous subsection, and as it will be detailed further in subsequent sections, conditions (I) to (V), as well as the assumption on the stochastic independence of processes X and W , ensure the absolutely continuity, in fact the mutual absolutely continuity, of several of the afore mentioned measures, and therefore the existence of the corresponding Radon-Nikodym derivatives. In particular,

$$\mu_{X, Y^r} \sim \mu_X \times \mu_\xi \text{ and } \mu_{Y^r} \sim \mu_\xi, \quad (14)$$

where, as usual, “ \sim ” denotes mutual absolutely continuity of the corresponding measures

and $\mu_X \times \mu_\xi$ the product measure in $(A_T \times C_T, \sigma(\mathcal{B}_{A_T} \times \mathcal{B}_T))$ obtained from μ_X and μ_ξ in (A_T, \mathcal{B}_{A_T}) and (C_T, \mathcal{B}_T) , respectively. From (14) it then follows that

$$\mu_{X,Y^r} \sim \mu_X \times \mu_{Y^r}$$

too. We denote the corresponding Radon-Nikodym derivatives by

$$\frac{d\mu_{X,Y^r}}{d[\mu_X \times \mu_\xi]}(f, g), \quad \frac{d\mu_{Y^r}}{d\mu_\xi}(g), \quad \text{and} \quad \frac{d\mu_{X,Y^r}}{d[\mu_X \times \mu_{Y^r}]}(f, g),$$

$f \in A_T$, $g \in C_T$. Note they are $\sigma(\mathcal{B}_{A_T} \times \mathcal{B}_T)$ -, \mathcal{B}_T -, and $\sigma(\mathcal{B}_{A_T} \times \mathcal{B}_T)$ -measurable functionals, respectively. For product measures, such as for example $\mu_X \times \mu_{Y^r}$, the differential $d[\mu_X \times \mu_{Y^r}]$ is sometimes written in the literature also as $d\mu_X d\mu_{Y^r}$.

In addition, for each $t \in [0, T]$, we denote by $\mu_{Y^r,t}$ and $\mu_{\xi,t}$ the measures the restricted processes $Y_{|[0,t]}^r \doteq (Y_s^r)_{s \in [0,t]}$ and $\xi_{|[0,t]} \doteq (\xi_s)_{s \in [0,t]}$ induce on (C_t, \mathcal{B}_t) , respectively, by

$$\frac{d\mu_{Y^r}}{d\mu_\xi}(t, g), \quad g \in C_t, \tag{15}$$

the corresponding Radon-Nikodym derivative, and similarly for all the other measures and processes above. In accordance with our previous notation, we omit t in expressions of the form (15) when $t = T$.

Finally, we denote by

$$\frac{d\mu_{Y^r}}{d\mu_\xi}(Y^r) \quad \text{and} \quad \frac{d\mu_{Y^r}}{d\mu_\xi}(t, Y^r)$$

the $\mathcal{F}_T^{Y^r}$ - and $\mathcal{F}_t^{Y^r}$ -measurable random variables, $t \in [0, T]$, obtained from the corresponding substitution of $g \in C_t$ in (15) by each sample path $(Y_s^r(\omega))_{s \in [0,t]}$, $\omega \in \Omega$, of the process $Y_{|[0,t]}^r$, and similarly for all other processes and measures above.

2.3 Input-Output Mutual Information

Let $\mathbb{R}^* \doteq \mathbb{R} \cup \{\pm\infty\}$, $\Theta(\Omega, \mathcal{F}, \mathbb{P})$ be the space of all \mathbb{R}^* -valued random variables θ on $(\Omega, \mathcal{F}, \mathbb{P})$, and $L^1(\Omega, \mathcal{F}, \mathbb{P})$ be the space of all $\theta \in \Theta$ having finite expectation, i.e.,

$$L^1(\Omega, \mathcal{F}, \mathbb{P}) \doteq \{\theta \in \Theta : \mathbb{E}[|\theta|] < \infty\},$$

with $\mathbb{E}[\cdot]$ denoting expectation w.r.t. \mathbb{P} and the usual measure theoretic convention $0[\pm\infty] = 0$.

We make the following definition involving the processes $X = (X_t)_{t \in [0, T]}$ and $Y^r = (Y_t^r)_{t \in [0, T]}$, $r \in \mathbb{R}_+$. Here, and throughout, logarithms are understood to be, without loss of generality, to the natural base e , with the convention $\log[0] = -\infty$.

Definition 2.1. *If for each $r \in \mathbb{R}_+$ the condition*

$$\log \left[\frac{d\mu_{X, Y^r}}{d[\mu_X \times \mu_{Y^r}]}(X, Y^r) \right] \in L^1(\Omega, \mathcal{F}, \mathbb{P}) \quad (16)$$

*is satisfied*¹⁰, *we define the input-output mutual information, $I : \mathbb{R}_+ \rightarrow \mathbb{R}$, by*

$$I(r) \doteq \mathbb{E} \left[\log \left[\frac{d\mu_{X, Y^r}}{d[\mu_X \times \mu_{Y^r}]}(X, Y^r) \right] \right]. \quad (17)$$

*In the same way, we define the instantaneous input-output mutual information, $I_i : [0, T] \times \mathbb{R}_+ \rightarrow \mathbb{R}$, by*¹¹

$$I_i(t, r) \doteq \mathbb{E} \left[\log \left[\frac{d\mu_{X, Y^r}}{d[\mu_X \times \mu_{Y^r}]}(t, X, Y^r) \right] \right].$$

Note that $I(r) = I_i(T, r)$ for each $r \in \mathbb{R}_+$. Note also that we may alternatively write $I(r)$ as

$$\int_{A_T \times C_T} \log \left[\frac{d\mu_{X, Y^r}}{d[\mu_X \times \mu_{Y^r}]}(f, g) \right] d[\mu_X \times \mu_{Y^r}](f, g),$$

$r \in \mathbb{R}_+$, and similarly for $I_i(t, r)$, $(t, r) \in [0, T] \times \mathbb{R}_+$.

Remark 2.3. *For a given input process X , changing the value of $r \in \mathbb{R}_+$ in (6) changes the output process Y^r , and thus changes the right hand side of (17) too. Therefore the notation $I(r)$, treating $r \in \mathbb{R}_+$ as the variable for a given input process X . The notation $I_i(t, r)$ obeys to the same reasoning. We find this notation more appealing than for example $I(X, Y^r)$ or $I_i(t, X, Y^r)$, specially in identifying the relevant variables to compute quantities such as*

$$\frac{d}{dr} I(r) \text{ and } \frac{\partial^2}{\partial t \partial r} I_i(t, r)$$

in subsequent sections.

Sufficient conditions for (16) to be satisfied will be discussed in subsequent sections.

It is easy to check that I and I_i are indeed non-negative-valued, i.e.,

$$I : \mathbb{R}_+ \rightarrow \mathbb{R}_+ \text{ and } I_i : [0, T] \times \mathbb{R}_+ \rightarrow \mathbb{R}_+.$$

¹⁰Note that, for each $r \in \mathbb{R}_+$, the left hand side of (16) is $\mathcal{F}_T^{X, Y^r} \doteq \sigma(\{X_t, Y_t^r : t \in [0, T]\})$ -measurable, therefore \mathcal{F} -measurable too, and hence an element of $\Theta(\Omega, \mathcal{F}, \mathbb{P})$.

¹¹Note condition (16) also implies the well definiteness of I_i .

Definition 2.1 is motivated from the classical definition of mutual information in the context of stochastic processes and stochastic systems [3, 34, 35], such as the AWGNC.

2.4 Minimum Mean-Square Errors

A central role will be played in all the results to be stated in the paper by the measurable non-anticipative functional

$$\phi : [0, T] \times A_T \times C_T \rightarrow \mathbb{R},$$

given by

$$\phi(t, f, g) \doteq \frac{F(t, f, g)}{G(t, g)}$$

for each $t \in [0, T]$, $f \in A_T$, and $g \in C_T$. Note from condition (III), equation (12), we have $G(\cdot, \cdot) \neq 0$, and therefore ϕ is well defined.

Remark 2.4. *From condition (V), equation (13), it follows that, for each $r \in \mathbb{R}_+$,*

$$\mathbb{E} [|F(t, X, Y^r)|] < \infty$$

for Lebesgue almost-every $t \in [0, T]$. Since also, from condition (III), equation (12), we have $|G(\cdot, \cdot)| \geq \sqrt{K} > 0$, we conclude that, for each $r \in \mathbb{R}_+$,

$$\mathbb{E} [| \phi(t, X, Y^r) |] < \infty,$$

for Lebesgue almost-every $t \in [0, T]$ too. Therefore, for any \mathcal{G} sub- σ -algebra of \mathcal{F} and each $r \in \mathbb{R}_+$ the conditional expectation

$$\mathbb{E} [\phi(t, X, Y^r) | \mathcal{G}]$$

is a well defined and finite \mathcal{G} -measurable random variable (in fact an element of $L^1(\Omega, \mathcal{G}, \mathbb{P}|_{\mathcal{G}})$ with $\mathbb{P}|_{\mathcal{G}}$ denoting the restriction of \mathbb{P} to \mathcal{G} [36]), for Lebesgue-almost every $t \in [0, T]$ as well. By defining it as $\alpha \in \mathbb{R}$ on the remaining Lebesgue-null subset of $[0, T]$, henceforth we treat it as a real-valued function in $t \in [0, T]$, for each $r \in \mathbb{R}_+$.

Having made the previous remark, we now introduce the following definition involving the above introduced functional ϕ , and the accompanying stochastic processes $(\phi(t, X, Y^r))_{t \in [0, T]}$, $r \in \mathbb{R}_+$.

Definition 2.2. *For each $r \in \mathbb{R}_+$ we define the causal minimum mean-square error*

(CMMSE) in estimating the stochastic process $\phi(\cdot, X, Y^r)$ at time $t \in [0, T]$ from the observations Y_s^r , $s \in [0, t]$, denoted $\text{cmmse}_\phi(t, r)$, by

$$\text{cmmse}_\phi(t, r) \doteq \mathbb{E} \left[\left(\phi(t, X, Y^r) - \mathbb{E} [\phi(t, X, Y^r) | \mathcal{F}_t^{Y^r}] \right)^2 \right].$$

Similarly, for each $r \in \mathbb{R}_+$ we define the non-causal minimum mean-square error (NCMMSE) in smoothing the stochastic process $\phi(\cdot, X, Y^r)$ at time $t \in [0, T]$ from the observations Y_s^r , $s \in [0, T]$, denoted $\text{ncmmse}_\phi(t, r)$, by

$$\text{ncmmse}_\phi(t, r) \doteq \mathbb{E} \left[\left(\phi(t, X, Y^r) - \mathbb{E} [\phi(t, X, Y^r) | \mathcal{F}_T^{Y^r}] \right)^2 \right].$$

In the same way, and slightly abusing notation, for each $r \in \mathbb{R}_+$, $t \in [0, T]$, and $s \in [0, t]$ we set

$$\text{ncmmse}_\phi(t, s, r) \doteq \mathbb{E} \left[\left(\phi(s, X, Y^r) - \mathbb{E} [\phi(s, X, Y^r) | \mathcal{F}_t^{Y^r}] \right)^2 \right],$$

the NCMMSE in smoothing the stochastic process $\phi(\cdot, X, Y^r)$ at time $s \in [0, t]$ from the observations Y_u^r , $u \in [0, t]$, with $t \in [0, T]$ and the convention of omitting the first of its three arguments when it equals T , i.e., $\text{ncmmse}_\phi(T, \cdot, \cdot) \equiv \text{ncmmse}_\phi(\cdot, \cdot)$. Note that the quantities just defined differ through the conditioning σ -algebras, and that $\text{ncmmse}_\phi(t, t, r) = \text{cmmse}_\phi(t, r)$ for each $t \in [0, T]$ and $r \in \mathbb{R}_+$.

Remark 2.5. From Remark 2.4 it follows that, for any \mathcal{G} sub- σ -algebra of \mathcal{F} and each $r \in \mathbb{R}_+$,

$$\left(\phi(t, X, Y^r) - \mathbb{E} [\phi(t, X, Y^r) | \mathcal{G}] \right)^2$$

is a well defined non-negative random variable for each $t \in [0, T]$, and therefore each of the three quantities introduced in Definition 2.2 above is a well defined $\mathbb{R}_+ \cup \{\infty\}$ -valued function of its corresponding arguments, clearly jointly measurable. Note the domain of $\text{ncmmse}_\phi(\cdot, \cdot, \cdot)$ is the set $\mathcal{D} \subseteq \mathbb{R}_+^3$ given by

$$\mathcal{D} \doteq \left\{ (t, s, r) \in \mathbb{R}_+^3 : t \in [0, T], s \in [0, t], r \in \mathbb{R}_+ \right\}.$$

3 Input-Output Mutual Information and CMMSE

In this section we provide a result relating input-output mutual information, I , and CMMSE, cmmse_ϕ , for the general dynamical input-output system (6). The result generalizes the classical Duncan's theorem for AWGNCs with or without feedback [2, 24]. It also provides a general condition guaranteeing the fulfilment of requirement (16) in Definition 2.1.

Theorem 3.1. *Assume that for each $r \in \mathbb{R}_+$ we have*

$$\int_0^T \text{cmmse}_\phi(t, r) dt < \infty.$$

Then for each $r \in \mathbb{R}_+$ we have

$$\log \left[\frac{d\mu_{X, Y^r}}{d[\mu_X \times \mu_{Y^r}]}(X, Y^r) \right] \in L^1(\Omega, \mathcal{F}, \mathbb{P}),$$

and the following relationship between I and cmmse_ϕ ,

$$I(r) = \frac{r}{2} \int_0^T \text{cmmse}_\phi(t, r) dt, \quad (18)$$

holds for each $r \in \mathbb{R}_+$ as well.

Before giving the proof of the theorem we make the following remark.

Remark 3.1. *Under a finite average power condition*

$$\int_0^T \mathbb{E} [F^2(t, X, Y^r)] dt < \infty, \quad r \in \mathbb{R}_+, \quad (19)$$

it follows that

$$\int_0^T \text{cmmse}_\phi(t, r) dt < \infty, \quad r \in \mathbb{R}_+.$$

Indeed, from (19) and condition (III), equation (12), we have

$$\int_0^T \mathbb{E} [\phi^2(t, X, Y^r)] dt < \infty, \quad r \in \mathbb{R}_+,$$

which implies, by standard properties of expectations and conditional expectations for finite second order moment random variables [36], and with $\eta_t^r \doteq \phi(t, X, Y^r)$ and $\tilde{\eta}_t^r \doteq$

$\mathbb{E}[\phi(t, X, Y^r) \mid \mathcal{F}_t^{Y^r}], r \in \mathbb{R}_+, t \in [0, T], \text{ that}$

$$\begin{aligned}
\int_0^T \text{cmmse}_\phi(t, r) dt &= \int_0^T \mathbb{E}[(\eta_t^r - \tilde{\eta}_t^r)^2] \\
&\leq \int_0^T \left(\sqrt{\mathbb{E}[(\eta_t^r)^2]} + \sqrt{\mathbb{E}[(\tilde{\eta}_t^r)^2]} \right)^2 dt \\
&\leq \int_0^T \left(2\sqrt{\mathbb{E}[(\eta_t^r)^2]} \right)^2 dt \\
&= 4 \int_0^T \mathbb{E}[\phi^2(t, X, Y^r)] dt \\
&< \infty, \quad r \in \mathbb{R}_+.
\end{aligned} \tag{20}$$

Relationship (18) had been previously proved in the especial case of AWGNCs (with or without feedback [2, 24]), and under condition (19).

Proof. Let $r \in \mathbb{R}_+$ be fixed throughout the proof. From conditions (I) to (V), the fact that the processes X and W are independent, and [25, Lemma 7.6, p.292] and [25, Lemma 7.7, p.293], we have that

$$\mu_{X, Y^r} \sim \mu_X \times \mu_\xi \text{ and } \mu_{Y^r} \sim \mu_\xi.$$

Therefore

$$\mu_{X, Y^r} \sim \mu_X \times \mu_{Y^r} \tag{21}$$

too, and, by [25, Theorem 7.23, p.289],

$$\frac{d\mu_{X, Y^r}}{d[\mu_X \times \mu_{Y^r}]}(X, Y^r) = \frac{d\mu_{X, \xi}}{d[\mu_X \times \mu_\xi]}(X, Y^r) \times \left(\frac{d\mu_{Y^r}}{d\mu_\xi}(X, Y^r) \right)^{-1}$$

with the right hand side of the above expression equaling

$$\exp \left\{ \sqrt{r} \int_0^T \frac{F(t, X, Y^r) - \overline{F}(t, Y^r)}{G(t, Y^r)} d\overline{W}_t^r \right\} \times \exp \left\{ -\frac{r}{2} \int_0^T \frac{(F(t, X, Y^r) - \overline{F}(t, Y^r))^2}{G^2(t, Y^r)} dt \right\},$$

\mathbb{P} -almost surely, where the non-anticipative functional \overline{F} satisfies, for Lebesgue-almost every $t \in [0, T]$,

$$\overline{F}(t, Y^r) = \mathbb{E}[F(t, X, Y^r) \mid \mathcal{F}_t^{Y^r}], \tag{22}$$

\mathbb{P} -almost surely as well, and where $\overline{W}^r = (\overline{W}_t^r, \mathcal{F}_t^{Y^r})_{t \in [0, T]}$ is a standard Brownian motion given by

$$\overline{W}_t^r \doteq \int_0^t \frac{dY_s^r - \sqrt{r} \overline{F}(s, Y^r) ds}{G(s, Y^r)}. \tag{23}$$

Thus, we find that

$$\log \left[\frac{d\mu_{X,Y^r}}{d[\mu_X \times \mu_{Y^r}]}(X, Y^r) \right] = \sqrt{r} \int_0^T \psi(t, X, Y^r) d\bar{W}_t^r - \frac{r}{2} \int_0^T \psi^2(t, X, Y^r) dt, \quad (24)$$

where

$$\psi(t, X, Y^r) \doteq \frac{F(t, X, Y^r) - \bar{F}(t, Y^r)}{G(t, Y^r)}.$$

Note that \bar{W}^r , even though it is obviously adapted to the filtration $(\mathcal{F}_t)_{t \in [0, T]}$ ($\supseteq (\mathcal{F}_t^{Y^r})_{t \in [0, T]}$), it is a martingale¹², and in fact a standard Brownian motion, w.r.t. the filtration $(\mathcal{F}_t^{Y^r})_{t \in [0, T]}$, but not w.r.t. the filtration $(\mathcal{F}_t)_{t \in [0, T]}$ to which the integrand $(\psi(t, X, Y^r))_{t \in [0, T]}$ is adapted¹³ (unless in the trivial case when X is not random but a fixed deterministic trajectory). \bar{W}^r is in fact a semimartingale relative to the filtration $(\mathcal{F}_t)_{t \in [0, T]}$, i.e., the sum of an $(\mathcal{F}_t)_{t \in [0, T]}$ -local martingale¹⁴ and an $(\mathcal{F}_t)_{t \in [0, T]}$ -adapted finite variation process¹⁵. Indeed, from (23) and (6) we find

$$\begin{aligned} d\bar{W}_t^r &= \frac{dY_s^r - \sqrt{r} \bar{F}(s, Y^r) ds}{G(s, Y^r)} \\ &= \sqrt{r} \psi(t, X, Y^r) dt + \frac{dY_s^r - \sqrt{r} F(s, X, Y^r) ds}{G(s, Y^r)} \\ &= \sqrt{r} \psi(t, X, Y^r) dt + dW_t \\ &\doteq dV_t^r + dM_t, \end{aligned} \quad (25)$$

with $(\mathcal{F}_t)_{t \in [0, T]}$ -local martingale (in fact martingale) component

$$M_t \doteq \int_0^t dW_s = W_t, \quad t \in [0, T], \quad (26)$$

and, from conditions (III) and (V), equations (12) and (13), respectively, with $(\mathcal{F}_t)_{t \in [0, T]}$ -adapted finite variation component process

$$V_t^r \doteq \sqrt{r} \int_0^t \psi(s, X, Y^r) ds, \quad t \in [0, T]. \quad (27)$$

¹²Recall a stochastic process $(Z_t)_{t \in [0, T]}$ is a martingale w.r.t. the filtration $(\mathcal{G}_t)_{t \in [0, T]}$ if it is adapted to that filtration and, for each $0 \leq s \leq t \leq T$, $\mathbb{E}[|Z_t|] < \infty$ and $\mathbb{E}[Z_t | \mathcal{G}_s] = Z_s$, \mathbb{P} -almost surely.

¹³As it will be discussed in Section 6, \bar{W}^r can be made into an $(\mathcal{F}_t)_{t \in [0, T]}$ -standard Brownian motion under an appropriate change of measure.

¹⁴Recall a stochastic process $(Z_t)_{t \in [0, T]}$ is a local martingale w.r.t. the filtration $(\mathcal{G}_t)_{t \in [0, T]}$ if there exists an increasing sequence of stopping times $\{T_n\}_{n=0}^\infty \subseteq [0, T]$ [4] such that each stopped process $(Z_{\min\{t, T_n\}})_{t \in [0, T]}$ is a martingale w.r.t. $(\mathcal{G}_t)_{t \in [0, T]}$.

¹⁵Recall a stochastic process $(Z_t)_{t \in [0, T]}$ is said to be of finite variation if, almost surely, all its paths or trajectories are finite variation functions on any subinterval of $[0, T]$ [37].

Therefore, from equations (24) to (27) we conclude

$$\log \left[\frac{d\mu_{X,Y^r}}{d[\mu_X \times \mu_{Y^r}]}(X, Y^r) \right] = \sqrt{r} \int_0^T \psi(t, X, Y^r) dW_t + \frac{r}{2} \int_0^T \psi^2(t, X, Y^r) dt. \quad (28)$$

Now, note for each $t \in [0, T]$ we have

$$\begin{aligned} \psi(t, X, Y^r) &= \frac{F(t, X, Y^r) - \bar{F}(t, Y^r)}{G(t, Y^r)} \\ &= \phi(t, X, Y^r) - \frac{\bar{F}(t, Y^r)}{G(t, Y^r)}, \end{aligned}$$

and, since $(G(t, Y^r))_{t \in [0, T]}$ is obviously adapted to the history $(\mathcal{F}_t^{Y^r})_{t \in [0, T]}$, from (22) we have, for Lebesgue almost-every $t \in [0, T]$,

$$\begin{aligned} \frac{\bar{F}(t, Y^r)}{G(t, Y^r)} &= \frac{\mathbb{E}[F(t, X, Y^r) | \mathcal{F}_t^{Y^r}]}{G(t, Y^r)} \\ &= \mathbb{E} \left[\frac{F(t, X, Y^r)}{G(t, Y^r)} | \mathcal{F}_t^{Y^r} \right] \\ &= \mathbb{E}[\phi(t, X, Y^r) | \mathcal{F}_t^{Y^r}], \end{aligned}$$

\mathbb{P} -almost surely. Thus, for Lebesgue almost-every $t \in [0, T]$ as well we have

$$\psi(t, X, Y^r) = \phi(t, X, Y^r) - \mathbb{E}[\phi(t, X, Y^r) | \mathcal{F}_t^{Y^r}], \quad (29)$$

\mathbb{P} -almost surely, hence, by Fubini's theorem [37], and since $\psi^2 \geq 0$,

$$\begin{aligned} \mathbb{E} \left[\int_0^T \psi^2(t, X, Y^r) dt \right] &= \int_0^T \mathbb{E}[\psi^2(t, X, Y^r)] dt \\ &= \int_0^T \text{cmmse}_\phi(t, r) dt < \infty, \end{aligned} \quad (30)$$

and therefore, since also $W = (W_t, \mathcal{F}_t)_{t \in [0, T]}$ is a standard Brownian motion and $(\psi(t, X, Y^r))_{t \in [0, T]}$ is adapted to the same filtration $(\mathcal{F}_t)_{t \in [0, T]}$ w.r.t. which W is a martingale, we conclude that

$$\left(\int_0^t \psi(s, X, Y^r) dW_s, \mathcal{F}_t \right)_{t \in [0, T]}$$

is a centered martingale [38], and then, in particular, that

$$\mathbb{E} \left[\int_0^T \psi(t, X, Y^r) dW_t \right] = 0. \quad (31)$$

Thus, from (28), (30) and (31) we conclude that

$$\log \left[\frac{d\mu_{X,Y^r}}{d[\mu_X \times \mu_{Y^r}]}(X, Y^r) \right] \in L^1(\Omega, \mathcal{F}, \mathbb{P}),$$

and that

$$\begin{aligned} I(r) &= \mathbb{E} \left[\sqrt{r} \int_0^T \psi(t, X, Y^r) dW_t + \frac{r}{2} \int_0^T \psi^2(t, X, Y^r) dt \right] \\ &= \frac{r}{2} \mathbb{E} \left[\int_0^T \psi^2(t, X, Y^r) dt \right] \\ &= \frac{r}{2} \int_0^T \mathbb{E} [\psi^2(t, X, Y^r)] dt. \end{aligned}$$

Equation (18) then follows from the previous expression in light of (29), proving the theorem. \square

Remark 3.2. *The assumption in Theorem 3.1 implies the Lebesgue almost-everywhere finiteness of $\text{cmmse}_\phi(\cdot, r)$ on $[0, T]$ for each $r \in \mathbb{R}_+$.*

Remark 3.3. *Under the assumption that*

$$\int_0^T \mathbb{E} [\phi^2(t, X, Y^r)] dt < \infty, \quad r \in \mathbb{R}_+,$$

it is also possible to give a proof of Theorem 3.1 by reducing system (6) to an AWGNC with feedback, which can be accomplished by using existence and uniqueness theorems for solutions of SDEs with general driving semimartingales and constructing appropriate implicitly defined measurable non-anticipative functionals, and then applying the known results for that case [24]. However, the proof given here, in addition to require a weaker assumption (see (20) in Remark 3.1), shows how explicit computations can be handled for the general case, which will be of use in subsequent sections.

As mentioned before, Theorem 3.1, which relates input-output mutual information (I) and CMMSE (cmmse_ϕ) for the general dynamical input-output system (6), generalizes the classical Duncan's theorem for AWGNCs with or without feedback [2, 24]. Indeed, for the AWGNC with feedback we have $G \equiv 1$, therefore $\phi \equiv F$, and hence equation (18) in Theorem 3.1 reduces to

$$I(r) = \frac{r}{2} \int_0^T \mathbb{E} \left[(F(t, X, Y^r) - \mathbb{E} [F(t, X, Y^r) | \mathcal{F}_t^{Y^r}])^2 \right] dt,$$

which in turn reduces for the AWGNC without feedback, where in addition $F(t, X, Y^r) =$

X_t for each $t \in [0, T]$ (see equation (7)), to

$$I(r) = \frac{r}{2} \int_0^T \mathbb{E} \left[(X_t - \mathbb{E} [X_t | \mathcal{F}_t^{Y^r}])^2 \right] dt,$$

with $r \in \mathbb{R}_+$ the channel SNR and

$$E \left[(X_t - \mathbb{E} [X_t | \mathcal{F}_t^{Y^r}])^2 \right]$$

the CMMSE in estimating X at time $t \in [0, T]$, X_t , from the observations Y_s^r , $s \in [0, t]$.

Note in the general case

$$\phi(\cdot, X, Y^r) = \frac{F(\cdot, X, Y^r)}{G(\cdot, Y^r)}$$

plays the role of X . (or $F(\cdot, X, Y^r)$) above.

4 On an appropriate Notion of SNR

In this section we discuss on conditions under which the parameter $r \in \mathbb{R}_+$ in (6) can be properly interpreted as an SNR parameter for such a general input-output system, in analogy with the AWGNC case [3] described by (7). These conditions will allow us to establish in the next section a useful and important relationship between input-output mutual information, $I(\cdot)$, and NCMSE, $\text{ncmmse}_\phi(\cdot, \cdot) \equiv \text{ncmmse}_\phi(T, \cdot, \cdot)$, for the general dynamical input-output system (6), generalizing a known relationship holding for AWGNCs [3].

Consider the AWGNC without feedback, described by equation (7), i.e.,

$$dY_t^r = \sqrt{r}X_t dt + dW_t, \quad t \in (0, T],$$

with X and Y^r the channel input and output, respectively, for a given fixed value of the parameter $r \in \mathbb{R}_+$. Here $F(\cdot, X, Y^r) = X$. and $G \equiv 1$. Then, the ratio between the instantaneous “signal component” power,

$$(\sqrt{r}F(\cdot, X, Y^r))^2 = rX^2,$$

and the instantaneous “noisy component” power,

$$G^2(\cdot, Y^r) \equiv 1,$$

is given by

$$\left(\frac{\sqrt{r}F(\cdot, X, Y^r)}{G(\cdot, Y^r)} \right)^2 = rX^2, \quad (32)$$

i.e., it is proportional to r for a given fixed input power level¹⁶. Therefore the interpretation of r as an SNR channel parameter.

The interpretation of r as an SNR channel parameter is not as straightforward as above for the standard AWGNC with feedback, described by the equation

$$dY_t^r = \sqrt{r}F(t, X, Y^r)dt + dW_t, \quad t \in (0, T].$$

Here, though $G \equiv 1$, we have

$$\left(\frac{\sqrt{r}F(\cdot, X, Y^r)}{G(\cdot, Y^r)} \right)^2 = rF^2(\cdot, X, Y^r),$$

and therefore r cannot be properly interpreted as an SNR channel parameter since, for instance, it may very well happen that an increment in r changes the corresponding output process Y^r in such a way that, say, $rF^2(\cdot, X, Y^r)$ becomes even smaller.

It should be noted that treating $F(\cdot, X, Y^r)$ as a “net channel input” (instead of X) does not solve the above difficulty since, except in trivial cases, it is not possible to maintain then a fixed reference input power level $F^2(\cdot, X, Y^r)$ while varying $r \in \mathbb{R}_+$.

Motivated from the above discussion, and interpreting the general input-output dynamical system (6) from a classical communication systems point of view, as described in Section 2, we now make the following definitions identifying general classes of systems, belonging to the setting given by (6), where a notion of SNR can be properly introduced.

Definition 4.1. *We say the dynamical input-output system (6) is a quasi-SNR-system if for any input process X as in Section 2 and corresponding family of associated output processes Y^r , $r \in \mathbb{R}_+$, the family of stochastic processes*

$$\left\{ \left(r\phi^2(t, X, Y^r) \right)_{t \in [0, T]} \right\}_{r \in \mathbb{R}_+}$$

is \mathbb{P} -almost surely non-decreasing, in the sense that for each $r_1, r_2 \in \mathbb{R}_+$ with $r_1 \leq r_2$,

$$r_1\phi^2(\cdot, X, Y^{r_1}) \leq r_2\phi^2(\cdot, X, Y^{r_2})$$

¹⁶Of course, and strictly speaking, what should be kept fixed in the random inputs case is the average input power, $\int_0^T \mathbb{E}[X_t^2]dt$, with the corresponding interpretation of (32) also in terms of average quantities. However, that does not alter the present discussion.

\mathbb{P} -almost surely, i.e., \mathbb{P} -almost surely as well,

$$\begin{aligned} r_1 \phi^2(t, X, Y^{r_1}) &= r_1 \frac{F^2(t, X, Y^{r_1})}{G^2(t, Y^{r_1})} \\ &\leq r_2 \frac{F^2(t, X, Y^{r_2})}{G(t, Y^{r_2})} = r_2 \phi^2(t, X, Y^{r_2}) \end{aligned}$$

for all $t \in [0, T]$.

Definition 4.2. We say the dynamical input-output system (6) is an SNR-system (with SRN parameter $r \in \mathbb{R}_+$) if there exists a measurable non-anticipative functional $\theta : [0, T] \times A_T \rightarrow \mathbb{R}_+$ such that

$$\phi^2(t, f, g) = \theta(t, f)$$

for all $t \in [0, T]$, $f \in A_T$, and $g \in C_T$. Note then, for any $r \in \mathbb{R}_+$ and X and Y^r related by (6),

$$r \phi^2(t, X, Y^r) = \frac{r F^2(t, X, Y^r)}{G^2(t, Y^r)} = r \theta(t, X)$$

for all $t \in [0, T]$.

Definition 4.3. We say the dynamical input-output system (6) is a strong-SNR-system (with SRN parameter $r \in \mathbb{R}_+$) if there exists a measurable non-anticipative functional $\eta : [0, T] \times A_T \rightarrow \mathbb{R}$ such that

$$\phi(t, f, g) = \eta(t, f)$$

for all $t \in [0, T]$, $f \in A_T$, and $g \in C_T$. Note then, for any $r \in \mathbb{R}_+$ and X and Y^r related by (6),

$$\sqrt{r} \phi(t, X, Y^r) = \frac{\sqrt{r} F(t, X, Y^r)}{G(t, Y^r)} = \sqrt{r} \eta(t, X)$$

for all $t \in [0, T]$.

We straightforwardly have that an strong-SNR-system is an SNR-system, and that an SNR-system is a quasi-SNR-system. Also, an SNR-system where the functionals F and G have the same sign, i.e., where

$$F(t, f, g)G(t, g) \geq 0$$

for all $t \in [0, T]$, $f \in A_T$, and $g \in C_T$, is clearly an strong-SNR-system. Indeed, since then $\phi \geq 0$, we can take for η in Definition 4.3

$$\eta = \sqrt{\theta},$$

with θ satisfying Definition 4.2.

Note when system (6) is an strong-SNR-system, say with measurable non-anticipative functional $\eta : [0, T] \times A_T \rightarrow \mathbb{R}$ in Definition 4.3, it may be written as

$$dY_t^r = \sqrt{r}\eta(t, X)G(t, Y^r)dt + G(t, Y^r)dW_t,$$

i.e., as

$$dY_t^r = G(t, Y^r) [\sqrt{r}\eta(t, X)dt + dW_t],$$

and therefore interpreted as a cascade of two systems: An AWGNC followed by a semimartingale SDE system, the output of the first acting as the semimartingale integrator in the second, i.e.,

$$Y_t^r = \int_0^t G(s, Y_s^r) dZ_s^r, \quad t \in [0, T], \quad (33)$$

with

$$dZ_s^r = \sqrt{r}\eta(s, X)dt + dW_s. \quad (34)$$

Alternatively, $G(\cdot, \cdot)$ can be looked at as a functional feedback modulator factor, modulating the AWGNC differential output dZ^r . Note however that from (34) we recognize $(Z_t^r)_{t \in [0, T]}$ as an unbounded variation semimartingale, and therefore the integral in (33) corresponds to a semimartingale stochastic integral and not to an standard pathwise Lebesgue-Stieltjes integral.

As it will be discussed in the next section, a quasi-SNR-system is not enough to have the relationship between input-output mutual information, $I(\cdot)$, and NCMMSE, $\text{nccmmse}_\phi(\cdot, \cdot) \equiv \text{nccmmse}_\phi(T, \cdot, \cdot)$, proved therein. However, for sake of completeness, we provide in the following lemma and its corollary sufficient conditions for system (6) to be a quasi-SNR-system. Conditions for system (6) to be an SNR-system or an strong-SNR-system are explicit in the corresponding definitions, since they only involve the structure of the functional ϕ .

Lemma 4.1. *Assume the measurable non-anticipative functionals F and G in (6) are such that*

$$F(t, f, g) = \overline{F}(t, f, g(t)) \text{ and } G(t, g) = \overline{G}(t, g(t)) \quad (35)$$

for all $t \in [0, T]$, $f \in A_T$, and $g \in C_T$, where \overline{F} and \overline{G} are measurable mappings from $[0, T] \times A_T \times \mathbb{R}$ and $[0, T] \times \mathbb{R}$ into \mathbb{R} , respectively. Let X be any input process as in Section 2, and assume that $\overline{F}(t, X, \cdot)$ satisfies the following Lipschitz condition, in a \mathbb{P} -almost surely basis,

$$|\overline{F}(t, X, y_1) - \overline{F}(t, X, y_2)|^2 \leq K_X |y_1 - y_2|^2 \quad (36)$$

for each $t \in [0, T]$ and all $y_1, y_2 \in \mathbb{R}$, where K_X is a bounded random variable. Then, for each $0 \leq r_1 \leq r_2 < \infty$, the corresponding output processes $Y^{r_1} = (Y_t^{r_1})_{t \in [0, T]}$ and

$Y^{r_2} = (Y_t^{r_2})_{t \in [0, T]}$ defined by (6) are such that

$$\mathbb{P}(Y_t^{r_1} \leq Y_t^{r_2}, t \in [0, T]) = 1.$$

Proof. First note that, since K_X in (36) is bounded, we may assume, without loss of generality, that it is a finite constant (for a given X). Then, by using Itô's formula [17] and proceeding by similar arguments as in the proof of [17, Proposition 2.18, p. 293], we find that, for each $t \in [0, T]$,

$$\mathbb{E}[\Delta_t^+] \leq K_X \int_0^t \mathbb{E}[\Delta_s^+] ds,$$

where, for each $t \in [0, T]$ as well,

$$\Delta_t^+ \doteq \max\{\Delta_t, 0\}$$

and

$$\Delta_t \doteq Y_t^{r_1} - Y_t^{r_2}.$$

Thus, from Gronwall's inequality [17] we conclude that

$$\mathbb{E}[\Delta_t^+] = 0$$

for each $t \in [0, T]$, and therefore

$$\mathbb{P}(Y_t^{r_1} \leq Y_t^{r_2}) = 1,$$

for each $t \in [0, T]$ too. The result now follows from the sample path continuity of the system outputs [17]. \square

Remark 4.1. Though as stated in Section 2 conditions (I) to (V) are assumed to hold throughout, the reader can verify that Lemma 4.1 holds indeed the same under just, in addition to (36) of course, condition (I) and condition (III), equation (10), which now takes the form

$$|\overline{G}(t, y_1) - \overline{G}(t, y_2)|^2 \leq D |y_1 - y_2|^2$$

for each $t \in [0, T]$ and all $y_1, y_2 \in \mathbb{R}$, with D a finite constant.

Corollary 4.1. Assume the same hypotheses as in Lemma 4.1, and that the functional ϕ , taking now the form $\phi = \frac{\overline{F}}{\overline{G}}$ with \overline{F} and \overline{G} defined by (35), is such that for each $t \in [0, T]$,

$f \in A_T$, and $g_1, g_2 \in C_T$ with $g_1(s) \leq g_2(s)$ for all $s \in [0, T]$,

$$\phi(t, f, g_1) = \frac{\overline{F}(t, f, g_1(t))}{\overline{G}(t, g_1(t))} \leq \frac{\overline{F}(t, f, g_2(t))}{\overline{G}(t, g_2(t))} = \phi(t, f, g_2).$$

Then the dynamical input-output system (6) is a quasi-SNR-system.

Proof. Let X be any input process as in Section 2, $0 \leq r_1 \leq r_2 < \infty$, and $Y^{r_1} = (Y_t^{r_1})_{t \in [0, T]}$ and $Y^{r_2} = (Y_t^{r_2})_{t \in [0, T]}$ be the corresponding output processes. Then, from Lemma 4.1 we have

$$\mathbb{P}(Y_t^{r_1} \leq Y_t^{r_2}, t \in [0, T]) = 1,$$

and therefore, \mathbb{P} -almost surely,

$$\begin{aligned} r_1 \phi^2(t, X, Y^{r_1}) &= r_1 \frac{\overline{F}^2(t, X, Y_t^{r_1})}{\overline{G}^2(t, Y_t^{r_1})} \\ &\leq r_2 \frac{\overline{F}^2(t, X, Y_t^{r_1})}{\overline{G}^2(t, Y_t^{r_1})} \\ &\leq r_2 \frac{\overline{F}^2(t, X, Y_t^{r_2})}{\overline{G}^2(t, Y_t^{r_2})} = r_2 \phi^2(t, X, Y^{r_2}) \end{aligned}$$

for all $t \in [0, T]$, proving the corollary. \square

5 Input-Output Mutual Information and NCMMSE

In this section we establish an also useful and interesting relationship, relating now input-output mutual information, $I(\cdot)$, and NCMMSE, $\text{nccmmse}_\phi(\cdot, \cdot) \equiv \text{nccmmse}_\phi(T, \cdot, \cdot)$, for the general dynamical input-output system (6), provided a sufficiently strong proper notion of SNR is taken into account, namely system (6) being an strong-SNR-system. Recall from the previous section that an SNR-system is also an strong-SNR-system when the functionals F and G have the same sign, i.e., when

$$F(t, f, g)G(t, g) \geq 0$$

for all $t \in [0, T]$, $f \in A_T$, and $g \in C_T$.

Consider once again the AWGNC, where $F(t, f, g) = f(t)$ for all $t \in [0, T]$, $f \in A_T$, and

$g \in C_T$, and where $G \equiv 1$, i.e., described by the equation

$$dY_t^r = \sqrt{r}X_t dt + dW_t, \quad t \in (0, T],$$

relating the channel input X and the channel output Y^r for each value of the parameter $r \in \mathbb{R}_+$. Then, provided¹⁷

$$\int_0^T \mathbb{E} [X_t^2] dt < \infty, \quad (37)$$

we have that [3] $I : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ is differentiable in \mathbb{R}_+ (from the right at the origin) and that the relationship

$$\frac{d}{dr}I(r) = \frac{1}{2} \int_0^T \text{ncmmse}_\phi(t, r) dt \quad (38)$$

holds for each $r \in \mathbb{R}_+$ (here $\phi(t, f, g) = f(t)$).

However, as pointed out in Guo et al. [3], relationship (38) does not hold true in the AWGNC with feedback, described by the equation

$$dY_t^r = \sqrt{r}F(t, X, Y^r)dt + dW_t, \quad t \in (0, T], \quad (39)$$

even if, in the terminology introduced in the previous section, system (39) is a quasi-SNR-system.

The following result establishes that relationship (38) does indeed hold for system (6), provided it is an strong-SNR-system.

Theorem 5.1. *Assume that system (6) is an strong-SNR-system and that the stochastic process $(\phi(t, X, Y^r))_{t \in [0, T]}$ has, for each $r \in \mathbb{R}_+$, finite average power, i.e.,*

$$\int_0^T \mathbb{E} [\phi^2(t, X, Y^r)] dt < \infty. \quad (40)$$

Then $I : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ is differentiable in \mathbb{R}_+ (from the right at the origin) and the following relationship between $I(\cdot)$ and $\text{ncmmse}_\phi(\cdot, \cdot)$,

$$\frac{d}{dr}I(r) = \frac{1}{2} \int_0^T \text{ncmmse}_\phi(t, r) dt, \quad (41)$$

holds for each $r \in \mathbb{R}_+$.

Before giving the proof of the theorem we make the following remark.

¹⁷As it can be read off from the proof of [3, Lemma 5], the finite average power condition (37) is now required.

Remark 5.1. Since in Theorem 5.1 system (6) is required to be an strong-SNR-system, say with measurable non-anticipative functional $\eta : [0, T] \times A_T \rightarrow \mathbb{R}$ in Definition 4.3, we have

$$dY_t^r = \sqrt{r}\eta(t, X)G(t, Y^r)dt + G(t, Y^r)dW_t,$$

and therefore condition (40) and relationship (41) take the form

$$\int_0^T \mathbb{E} [\eta^2(t, X)] dt < \infty$$

and

$$\frac{d}{dr}I(r) = \frac{1}{2} \int_0^T \mathbb{E} \left[\left(\eta(t, X) - \mathbb{E} [\eta(t, X) | \mathcal{F}_T^{Y^r}] \right)^2 \right] dt,$$

respectively.

Proof. As in Remark 5.1 above, for each $r \in \mathbb{R}_+$ we may write

$$dY_t^r = \sqrt{r}\eta(t, X)G(t, Y^r)dt + G(t, Y^r)dW_t,$$

and therefore, since from condition (III), equation (12), we have that $G \neq 0$, we may as well write

$$\frac{dY_t^r}{G(t, Y^r)} = \sqrt{r}\eta(t, X)dt + dW_t.$$

Define, for each $r \in \mathbb{R}_+$, the process $Z^r = (Z_t^r)_{t \in [0, T]}$ by

$$dZ_t^r \doteq \frac{dY_t^r}{G(t, Y^r)}, \quad Z_0^r = 0,$$

i.e., by

$$Z_t^r \doteq \int_0^t \frac{dY_s^r}{G(s, Y^r)}, \quad t \in [0, T]. \quad (42)$$

Note process Z^r has trajectories, the same as Y^r , in the measurable space (C_T, \mathcal{B}_T) . We may look at Z^r as being the output of the system

$$dZ_t^r = \sqrt{r}\eta(t, X)dt + dW_t, \quad (43)$$

corresponding to the input X and parameter r . System (43) is nothing but an AWGNC. Now, since

$$\int_0^T \mathbb{E} [\eta^2(t, X)] dt = \int_0^T \mathbb{E} [\phi^2(t, X, Y^r)] dt < \infty,$$

from Theorem 3.1 applied to system (43) (see Remark 3.1) we obtain

$$\log \left[\frac{d\mu_{X,Z^r}}{d[\mu_X \times \mu_{Z^r}]}(X, Z^r) \right] \in L^1(\Omega, \mathcal{F}, \mathbb{P}),$$

for each $r \in \mathbb{R}_+$, and

$$\mathbb{E} \left[\log \left[\frac{d\mu_{X,Z^r}}{d[\mu_X \times \mu_{Z^r}]}(X, Z^r) \right] \right] = \frac{r}{2} \int_0^T \mathbb{E} \left[(\eta(t, X) - \mathbb{E}[\eta(t, X) | \mathcal{F}_t^{Z^r}])^2 \right] dt,$$

for each $r \in \mathbb{R}_+$ as well. Moreover [3], the previous expression is differentiable in $r \in \mathbb{R}_+$ (from the right at the origin) and

$$\frac{d}{dr} \mathbb{E} \left[\log \left[\frac{d\mu_{X,Z^r}}{d[\mu_X \times \mu_{Z^r}]}(X, Z^r) \right] \right] = \frac{1}{2} \int_0^T \mathbb{E} \left[(\eta(t, X) - \mathbb{E}[\eta(t, X) | \mathcal{F}_T^{Z^r}])^2 \right] dt \quad (44)$$

holds for each $r \in \mathbb{R}_+$, with $\mathcal{F}_T^{Z^r} = \mathcal{F}_{t=T}^{Z^r}$ and

$$\mathcal{F}_t^{Z^r} \doteq \sigma(\{Z_s^r : s \in [0, t]\}),$$

the history of Z^r up to time $t \in [0, T]$. But, from the definition of process Z^r in (42), it is clear that

$$\mathcal{F}_t^{Z^r} \subseteq \mathcal{F}_t^{Y^r}$$

for each $t \in [0, T]$. In addition, we may also rewrite (42) as

$$Y_t^r = \int_0^t G(s, Y_s^r) dZ_s^r, \quad t \in [0, T],$$

and regard the previous expression as an SDE being satisfied by $Y^r = (Y_t^r)_{t \in [0, T]}$, with “driving” semimartingale $Z^r = (Z_t^r)_{t \in [0, T]}$ given by (43). Then, by the existence and uniqueness theorem [4, Theorem 7, p.253], and from the Lipschitz continuity requirement in condition (III), equation (10), we conclude that

$$\mathcal{F}_t^{Y^r} \subseteq \mathcal{F}_t^{Z^r}$$

for each $t \in [0, T]$, and therefore

$$\mathcal{F}_t^{Z^r} = \mathcal{F}_t^{Y^r},$$

for each $t \in [0, T]$ as well. Thus, from [25, Lemma 4.9, p.114] we conclude the existence, for each $r \in \mathbb{R}_+$, of measurable non-anticipative functionals a_r and b_r , from $[0, T] \times C_T$ into \mathbb{R} , such that

$$Z_t^r(\omega) = a_r(t, Y^r(\omega)) \text{ and } Y_t^r(\omega) = b_r(t, Z^r(\omega))$$

for $(\lambda \times \mathbb{P})$ -almost every $(t, \omega) \in [0, T] \times \Omega$, with λ denoting Lebesgue measure in $[0, T]$ and $\lambda \times \mathbb{P}$ the product measure of λ and \mathbb{P} . Hence, we may replace in (44) all occurrences of Z^r by Y^r to obtain

$$\frac{d}{dr} \mathbb{E} \left[\log \left[\frac{d\mu_{X, Y^r}}{d[\mu_X \times \mu_{Y^r}]}(X, Y^r) \right] \right] = \frac{1}{2} \int_0^T \mathbb{E} \left[(\eta(t, X) - \mathbb{E}[\eta(t, X) | \mathcal{F}_T^{Y^r}])^2 \right] dt,$$

giving us (41) (see Remark 5.1) and thus proving the theorem. \square

We have the following corollary to Theorem 5.1, generalizing the corresponding result for AWGNCs [3].

Corollary 5.1. *Under the same assumptions as in Theorem 5.1, for each $r \in (0, \infty)$ we have*

$$\overline{\text{cmmse}}_\phi(r) = \frac{1}{r} \int_0^r \overline{\text{ncmmse}}_\phi(u) du,$$

with

$$\overline{\text{cmmse}}_\phi(\cdot) \doteq \frac{1}{T} \int_0^T \text{cmmse}_\phi(t, \cdot) dt$$

and

$$\overline{\text{ncmmse}}_\phi(\cdot) \doteq \frac{1}{T} \int_0^T \text{ncmmse}_\phi(t, \cdot) dt$$

the time-averaged CMMSE and NCMMSE over $[0, T]$, respectively.

Proof. The result follows directly from Remark 3.1 and Theorems 3.1 and 5.1. \square

6 Dynamical Relationships

It is apparent from the previous sections that the results already provided have time-instantaneous counterparts, and in particular that we have consistency in that

$$\frac{\partial}{\partial t} I_i(t, r) = \frac{r}{2} \text{cmmse}_\phi(t, r)$$

and

$$\frac{\partial}{\partial r} I_i(t, r) = \frac{1}{2} \int_0^t \text{ncmmse}_\phi(t, s, r) ds.$$

Remark 6.1 and Theorem 6.1 below show that is indeed true. This fact brings dynamical relationships into the picture allowing to write general integro-partial differential equations,

also given in this section, characterizing instantaneous input-output mutual information and MMSEs.

Remark 6.1. *Consider the condition*

$$\mathbb{P} \left(\int_0^T \psi^2(t, X, Y^r) dt < \infty \right) = 1, \quad r \in \mathbb{R}_+,$$

with

$$\psi(s, X, Y^r) \doteq \phi(s, X, Y^r) - \mathbb{E} [\phi(s, X, Y^r) | \mathcal{F}_s^{Y^r}],$$

which is implied by conditions (IV) and (V), and define the process $M^r = (M_t^r, \mathcal{F}_t)_{t \in [0, T]}$ by

$$\begin{aligned} M_t^r &\doteq \exp \left\{ -\sqrt{r} \int_0^t \psi(s, X, Y^r) dW_s \right\} \\ &\quad \times \exp \left\{ -\frac{r}{2} \int_0^t \psi^2(s, X, Y^r) ds \right\}. \end{aligned}$$

Note from the proof of Theorem 3.1 we have

$$\frac{d[\mu_X \times \mu_{Y^r}]}{d\mu_{X, Y^r}}(X, Y^r) = M_T^r,$$

i.e. $(\mu_{X, Y^r} \sim \mu_X \times \mu_{Y^r})$,

$$\frac{d\mu_{X, Y^r}}{d[\mu_X \times \mu_{Y^r}]}(X, Y^r) = (M_T^r)^{-1},$$

\mathbb{P} -almost surely. Also note that $(M_t^r, \mathcal{F}_t)_{t \in [0, T]}$ is a (strictly positive) supermartingale¹⁸ [25], and, since furthermore

$$\mathbb{E} [M_T^r] = \mathbb{E} \left[\frac{d[\mu_X \times \mu_{Y^r}]}{d\mu_{X, Y^r}}(X, Y^r) \right] = 1 = \mathbb{E} [M_0^r],$$

we have that $(M_t^r, \mathcal{F}_t)_{t \in [0, T]}$ is in fact a martingale [25]. Hence, for each $t \in [0, T]$ we have

$$\mathbb{E} [M_T^r | \mathcal{F}_t] = M_t^r, \quad \mathbb{P}\text{-almost surely, } \mathbb{E} [M_t^r] = 1,$$

and therefore the consistency property [25]

$$\frac{d[\mu_X \times \mu_{Y^r}]}{d\mu_{X, Y^r}}(t, X, Y^r) = M_t^r,$$

¹⁸Recall a stochastic process $(Z_t)_{t \in [0, T]}$ is a supermartingale w.r.t. the filtration $(\mathcal{G}_t)_{t \in [0, T]}$ if it is adapted to that filtration and, for each $0 \leq s \leq t \leq T$, $\mathbb{E}[|Z_t|] < \infty$ and $\mathbb{E}[Z_t | \mathcal{G}_s] \leq Z_s$, \mathbb{P} -almost surely.

\mathbb{P} -almost surely as well, $t \in [0, T]$. Equivalently, in terms of $((M_t^r)^{-1}, \mathcal{F}_t)_{t \in [0, T]}$, with $\mathbb{Q}^r (\sim \mathbb{P})$ the probability measure on (Ω, \mathcal{F}_T) given by

$$\mathbb{Q}^r(A) \doteq \int_A M_T^r d\mathbb{P}, \quad A \in \mathcal{F}_T,$$

and with $\mathbb{E}_{\mathbb{Q}^r}[\cdot|\cdot]$ (resp., $\mathbb{E}_{\mathbb{Q}^r}[\cdot]$) denoting conditional expectation (resp., expectation) on $(\Omega, \mathcal{F}_T, \mathbb{Q}^r)$, for each $t \in [0, T]$ we have [17]

$$\begin{aligned} \mathbb{E}_{\mathbb{Q}^r} [(M_T^r)^{-1} | \mathcal{F}_t] &= (M_t^r)^{-1} \mathbb{E} [(M_t^r)^{-1} M_t^r | \mathcal{F}_t] \\ &= (M_t^r)^{-1}, \end{aligned}$$

\mathbb{Q}^r -almost surely, therefore we have that $((M_t^r)^{-1}, \mathcal{F}_t)_{t \in [0, T]}$ is a martingale on $(\Omega, \mathcal{F}_T, \mathbb{Q}^r)$, with

$$\begin{aligned} \mathbb{E}_{\mathbb{Q}^r} [(M_t^r)^{-1}] &= \mathbb{E}_{\mathbb{Q}^r} [(M_T^r)^{-1}] \\ &= \int_{\Omega} \frac{d\mathbb{P}}{d\mathbb{Q}^r} d\mathbb{Q}^r = \mathbb{P}(\Omega) = 1, \quad t \in [0, T], \end{aligned}$$

and, as before, for each $t \in [0, T]$ we thus have the consistency property [25]

$$\frac{d\mu_{X, Y^r}}{d[\mu_X \times \mu_{Y^r}]}(t, X, Y^r) = (M_t^r)^{-1},$$

\mathbb{Q}^r -almost surely, hence \mathbb{P} -almost surely too since in particular \mathbb{P} is absolutely continuous w.r.t. \mathbb{Q}^r . Alternatively, and in connection with the proof of Theorem 3.1, note since M^r is an $(\mathcal{F}_t)_{t \in [0, T]}$ -martingale we have with \overline{W}^r as defined there, and from Girsanov's theorem [4, 17], that¹⁹ $(\overline{W}_t^r, \mathcal{F}_t)_{t \in [0, T]}$ is a standard Brownian motion on $(\Omega, \mathcal{F}_T, \mathbb{Q}^r)$, and therefore, by the same arguments as before, the process

$$\begin{aligned} (M_t^r)^{-1} &= \exp \left\{ \sqrt{r} \int_0^t \psi(s, X, Y^r) d\overline{W}_s^r \right\} \\ &\quad \times \exp \left\{ -\frac{r}{2} \int_0^t \psi^2(s, X, Y^r) ds \right\}, \quad t \in [0, T], \end{aligned} \tag{45}$$

is an $(\mathcal{F}_t)_{t \in [0, T]}$ -martingale on $(\Omega, \mathcal{F}_T, \mathbb{Q}^r)$. Indeed, (45) follows from the proof of Theorem 3.1, and, since

$$\mathbb{Q}^r \left(\int_0^T \psi^2(t, X, Y^r) dt < \infty \right) = 1,$$

the right hand side of (45) is an $(\mathcal{F}_t)_{t \in [0, T]}$ -supermartingale on $(\Omega, \mathcal{F}_T, \mathbb{Q}^r)$ having constant

¹⁹Note that, rather than considering the tuple $(\overline{W}_t^r, \mathcal{F}_t^{Y^r})_{t \in [0, T]}$ on the space $(\Omega, \mathcal{F}, \mathbb{P})$ as in the proof of Theorem 3.1, we now consider the tuple $(\overline{W}_t^r, \mathcal{F}_t)_{t \in [0, T]}$ on the space $(\Omega, \mathcal{F}_T, \mathbb{Q}^r)$.

expectation, hence a martingale [25].

Having stated the previous remark, we now give the main results of this section.

Theorem 6.1. *Assume that for each $r \in \mathbb{R}_+$ we have*

$$\int_0^T \text{cmmse}_\phi(t, r) dt < \infty.$$

Then, for each $r \in \mathbb{R}_+$ as well, $I_i(\cdot, r) : [0, T] \rightarrow \mathbb{R}_+$ is Lebesgue-almost everywhere differentiable in $[0, T]$ and, at each point $t \in [0, T]$ where this is so, we have

$$\frac{\partial}{\partial t} I_i(t, r) = \frac{r}{2} \text{cmmse}_\phi(t, r).$$

Moreover, if in addition system (6) is an strong-SNR-system and for each $r \in \mathbb{R}_+$ we have

$$\int_0^T \mathbb{E} [\phi^2(t, X, Y^r)] dt < \infty, \quad (46)$$

then, for each $t \in [0, T]$, $I_i(t, \cdot) : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ is differentiable in \mathbb{R}_+ (from the right at the origin) with derivative given by

$$\frac{\partial}{\partial r} I_i(t, r) = \frac{1}{2} \int_0^t \text{ncmmse}_\phi(t, s, r) ds \quad (47)$$

for each $r \in \mathbb{R}_+$.

Before giving the proof of the theorem we make the following remark.

Remark 6.2. *As in remark 5.1 in the previous section, note that condition (46) and relationship (47) take the form*

$$\int_0^T \mathbb{E} [\eta^2(t, X)] dt < \infty$$

and

$$\frac{\partial}{\partial r} I_i(t, r) = \frac{1}{2} \int_0^t \mathbb{E} \left[(\eta(s, X) - \mathbb{E} [\eta(s, X) | \mathcal{F}_t^{Y^r}])^2 \right] ds,$$

respectively, with system (6) an strong-SNR-system satisfying Definition 4.3 with measurable non-anticipative functional η .

Proof. Let $r \in \mathbb{R}_+$. From Remark 6.1 we have, for each $t \in [0, T]$,

$$\frac{d\mu_{X,Y^r}}{d[\mu_X \times \mu_{Y^r}]}(X, Y^r, t) = \exp \left\{ \sqrt{r} \int_0^t \psi(s, X, Y^r) dW_s \right\} \times \exp \left\{ \frac{r}{2} \int_0^t \psi^2(s, X, Y^r) ds \right\},$$

\mathbb{P} -almost surely, from where, and proceeding by the same arguments as in the proof of Theorem 3.1,

$$I_i(t, r) = \frac{r}{2} \int_0^t \text{cmmse}_\phi(s, r) ds.$$

The first part of the theorem then follows. The second part of the theorem also follows from the previous relationship by applying it to an AWGNC as in the the proof of Theorem 5.1, when system (6) is an strong-SNR-system, and proceeding by the same arguments considered therein. The theorem is then proved. \square

We have the following two corollaries to Theorem 6.1.

Corollary 6.1. *Assume that system (6) is an strong-SNR-system and that for each $r \in \mathbb{R}_+$ we have*

$$\int_0^T \mathbb{E} [\phi^2(t, X, Y^r)] dt < \infty.$$

Then, for each $r \in (0, \infty)$ and $t \in (0, T]$ we have

$$\overline{\text{cmmse}}_\phi(t, r) = \frac{1}{r} \int_0^r \overline{\text{ncmmse}}_\phi(t, u) du, \quad (48)$$

with

$$\overline{\text{cmmse}}_\phi(t, \cdot) \doteq \frac{1}{t} \int_0^t \text{cmmse}_\phi(s, \cdot) ds$$

and

$$\overline{\text{ncmmse}}_\phi(t, \cdot) \doteq \frac{1}{t} \int_0^t \text{ncmmse}_\phi(t, s, \cdot) ds$$

the time-averaged CMMSE and NCMMSE over $[0, t]$, respectively.

Proof. The result follows directly from Remark 3.1 and Theorem 6.1. \square

For the next corollary, denote as usual by $\mathcal{C}^k(A)$ the space of functions $h : A \rightarrow \mathbb{R}$ with continuous k -th order partial derivatives in $A \subseteq \mathbb{R}^n$. Partial derivatives at a boundary point are understood to be taken from the right or from the left, accordingly. In the same way, denote by $\mathcal{C}^0(A)$ the space of functions $h : A \rightarrow \mathbb{R}$ continuous in $A \subseteq \mathbb{R}^n$, with an analogous convention than before at boundary points.

Corollary 6.2. *Assume the same hypotheses as in Corollary 6.1. Assume furthermore that $\text{cmmse}_\phi(\cdot, \cdot) \in \mathcal{C}^1([0, T] \times \mathbb{R}_+)$ and that $\text{ncmmse}_\phi(\cdot, \cdot, \cdot)$ is differentiable w.r.t. its first and third arguments, $t \in [0, T]$ and $r \in \mathbb{R}_+$ respectively, with*

$$\frac{\partial}{\partial t} \text{ncmmse}_\phi(\cdot, \cdot, \cdot) \text{ and } \frac{\partial}{\partial r} \text{ncmmse}_\phi(\cdot, \cdot, \cdot)$$

both belonging to²⁰ $\mathcal{C}^0(\mathcal{D})$. Then,

$$I_i(\cdot, \cdot) \in \mathcal{C}^2([0, T] \times \mathbb{R}_+)$$

with second-order partial derivatives given, for each $(t, r) \in [0, T] \times \mathbb{R}_+$, by

$$\begin{aligned} 2 \frac{\partial^2}{\partial t^2} I_i(t, r) &= r \frac{\partial}{\partial t} \text{cmmse}_\phi(t, r), \\ 2 \frac{\partial^2}{\partial r^2} I_i(t, r) &= \int_0^t \frac{\partial}{\partial r} \text{ncmmse}_\phi(t, s, r) ds, \\ 2 \frac{\partial^2}{\partial t \partial r} I_i(t, r) &= \int_0^t \frac{\partial}{\partial t} \text{ncmmse}_\phi(t, s, r) ds + \text{cmmse}_\phi(t, r), \end{aligned}$$

and

$$2 \frac{\partial^2}{\partial r \partial t} I_i(t, r) = r \frac{\partial}{\partial r} \text{cmmse}_\phi(t, r) + \text{cmmse}_\phi(t, r).$$

In particular,

$$r \frac{\partial}{\partial r} \text{cmmse}_\phi(t, r) = \int_0^t \frac{\partial}{\partial t} \text{ncmmse}_\phi(t, s, r) ds, \quad (49)$$

for each $(t, r) \in [0, T] \times \mathbb{R}_+$ as well.

Before giving the proof of the corollary we make the following remarks.

Remark 6.3. *It is easy to see that, under the assumptions of Corollary 6.2, equation (49) can also be obtained from (48) by multiplying both sides of (48) by rt and then taking the derivative $\frac{\partial^2}{\partial r \partial t}$ to the resulting equation (using Leibniz's rule as before), and therefore relationship (48) corresponds to an integrated version of (49).*

Remark 6.4. *The smoothness requirements on cmmse_ϕ and ncmmse_ϕ in Corollary 6.2 can be guaranteed under appropriate corresponding smoothness requirements on the coefficients F and G , for certain input-trajectory spaces A_T and structures of F and G [25].*

²⁰Recall $\mathcal{D} \doteq \{(t, s, r) \in \mathbb{R}_+^3 : t \in [0, T], s \in [0, t], r \in \mathbb{R}_+\}$.

Proof. That $I_i(\cdot, \cdot) \in \mathcal{C}^2([0, T] \times \mathbb{R}_+)$ with the corresponding given expressions for the second order partial derivatives follows directly from the assumptions, Remark 3.1 and the expressions for the first order partial derivatives in Theorem 6.1, and the use of Leibniz's rule for the differentiation of integrals [39] along with the fact that $\text{ncmmse}_\phi(t, t, r) = \text{cmmse}_\phi(t, r)$ for each $t \in [0, T]$ and $r \in \mathbb{R}_+$. The last claim of the corollary, equation (49), follows from the fact that $I_i(\cdot, \cdot) \in \mathcal{C}^2([0, T] \times \mathbb{R}_+)$, since then we have

$$\frac{\partial^2}{\partial t \partial r} I_i(t, r) = \frac{\partial^2}{\partial r \partial t} I_i(t, r)$$

for each $(t, r) \in [0, T] \times \mathbb{R}_+$. □

7 Further Extensions and Results

As mentioned in Section 1, it is possible to give $n(> 1)$ -dimensional counterparts of all the results established in the paper, where system (6) takes the form

$$Y_t^r = \sqrt{r} \int_0^t F(s, X, Y^r) ds + \int_0^t G(s, Y^r) dW_s, \quad t \geq 0, \quad (50)$$

with $r \in \mathbb{R}_+$, $X = (X^i)_{i=1}^n$ and $Y^r = (Y^{r,i})_{i=1}^n$ the \mathbb{R}^n -valued²¹ input and output processes, respectively, $W = (W^i)_{i=1}^n$ an \mathbb{R}^n -valued standard Brownian motion independent of X , $F(\cdot, \cdot, \cdot) = (F^i(\cdot, \cdot, \cdot))_{i=1}^n$ an n -dimensional vector of \mathbb{R} -valued measurable non-anticipative functionals, and $G(\cdot, \cdot) = (G^{i,j}(\cdot, \cdot))_{i,j=1}^n$ an $n \times n$ matrix of \mathbb{R} -valued measurable non-anticipative functionals as well. The corresponding Radon-Nikodym derivatives, and consequently the input-output mutual information, are then defined in terms of the measures the different processes involved induce in the corresponding multi-dimensional space, as for example (C_T^n, \mathcal{B}_T^n) , the space of \mathbb{R}^n -valued continuous function in $[0, T]$ equipped with the corresponding σ -algebra of cylinder sets, similarly to the considered case $n = 1$. The required changes in the statement of the corresponding n -dimensional results are straightforward, with the functional ϕ taking now the form

$$\phi(t, f, g) = (\phi_i(t, f, g))_{i=1}^n = [G(t, g)]^{-1} F(t, f, g), \quad (51)$$

$t \in [0, T]$, $f \in A_T^n$, $g \in C_T^n$, and with condition (III), equation (12), now interpreted as the requirement of

$$H(\cdot, \cdot) = (H_{i,j}(\cdot, \cdot))_{i,j=1}^n \doteq G(\cdot, \cdot) [G(\cdot, \cdot)]^*,$$

²¹All vectors in \mathbb{R}^n or vector-valued processes should be envisioned as column vectors.

with $[\cdot]^*$ denoting the transpose of the corresponding matrix (or vector), being uniformly elliptic [40], i.e., such that there exists $\delta \in (0, \infty)$ with

$$\sum_{i,j=1}^n H_{i,j}(t, g) \gamma_i \gamma_j \geq \delta \|\gamma\|^2$$

for all $t \in [0, T]$, $g \in C_T^n$, and $\gamma = (\gamma_i)_{i=1}^n \in \mathbb{R}^n$, where $\|\cdot\|$ denotes the usual Euclidian norm in \mathbb{R}^n , i.e., $\|y\| \doteq y^* y$ for each $y \in \mathbb{R}^n$. Note the uniform ellipticity of H in particular implies the invertibility of G . All other requirements in Section 2, Subsection 2.1, on the functionals F and G , or on processes such as $(F(t, X, Y^r))_{t \in [0, T]}$, are understood to hold in the n -dimensional setting in a componentwise (or elementwise, in case of matrices) fashion. Equivalently, they can be written in terms of the Euclidian norm $\|\cdot\|$, the 1-norm $\|y\|_1 \doteq \sum_{i=1}^n |y_i|$, $y = (y_i)_{i=1}^n \in \mathbb{R}^n$, or the Frobenious matrix norm $\|\cdot\|_F$, given by

$$\|A\|_F^2 \doteq \sum_{i,j=1}^n A_{i,j}^2, \quad A = (A_{i,j})_{i,j=1}^n,$$

accordingly. Similarly, conditions such as

$$\int_0^T \mathbb{E} [\phi^2(t, X, Y^r)] dt < \infty$$

are also interpreted as to holding in a componentwise fashion or, equivalently, in terms of the Euclidian norm $\|\cdot\|$,

$$\begin{aligned} \int_0^T \mathbb{E} [\|\phi(t, X, Y^r)\|^2] dt &= \sum_{i=1}^n \int_0^T \mathbb{E} [\phi_i^2(t, X, Y^r)] dt \\ &< \infty. \end{aligned}$$

In the same way, MMSEs are written in terms of the Euclidian norm, like for instance

$$\begin{aligned} \text{cmmse}_\phi(t, r) &= \mathbb{E} \left[\left\| \phi(t, X, Y^r) - \mathbb{E} [\phi(t, X, Y^r) | \mathcal{F}_t^{Y^r}] \right\|^2 \right] \\ &= \sum_{i=1}^n \mathbb{E} \left[\left(\phi_i(t, X, Y^r) - \mathbb{E} [\phi_i(t, X, Y^r) | \mathcal{F}_t^{Y^r}] \right)^2 \right], \end{aligned}$$

$t \in [0, T]$, $r \in \mathbb{R}_+$. The analogous definitions for system (50) to be a quasi-SNR-system, an SNR-system, or an strong-SNR-System are also straightforward from Section 4, with ϕ given by (51) and the obvious replacement of ϕ^2 by $\|\phi\|^2$.

It is also possible to consider system (50) in the case when $r = (r_i)_{i=1}^n \in \mathbb{R}^n$, with \sqrt{r}

in (50) replaced by the diagonal matrix

$$\text{diag}(\sqrt{r_1}, \dots, \sqrt{r_n}),$$

and to give relationships involving not only time derivatives of the input-output mutual information, but also, the same as for the AWGNC case, derivatives w.r.t. each component r_i of r . We do not give the details since, in light of the results already stated in the paper, this extension follows by the same line of arguments as in the AWGNC case [3].

Finally, and again in light of the results already stated in the paper, it is also possible to study the asymptotics of input-output mutual information and MMSEs, writing analogous expressions as in the AWGNC case for high and low values of $r \in \mathbb{R}_+$, and to find representations of other information measures such as entropy and divergence in terms of pure estimation-theoretic quantities, also analogous to the AWGNC case [3]. The details are left to the reader.

8 Conclusion

In this paper we have considered a general stochastic input-output dynamical system, covering a wide range of stochastic system models appearing in engineering applications. In such general setting, we have established important relationships linking information and estimation theoretic quantities. In particular, precise equations revealing the connection between input-output mutual information and minimum mean causal and non-causal square errors were found for this setting, corresponding to analogous of previously known results in the context of additive Gaussian noise communication channels. Furthermore, they were stated here in this broader setting not only in terms of time-averaged quantities, but also their time-instantaneous, dynamical counterparts were presented. In extending those relationships we have also identified conditions for a signal-to-noise ratio parameter to be meaningful, and characterized in those terms different system model classes.

We believe the results presented in the paper will find interesting future applications in several engineering fields, as they evidence that the deep connection between information theory and estimation theory goes beyond communication systems, encompassing indeed a whole range of dynamical systems of great use and interest in the stochastic modelling community.

References

- [1] R. R. Mazumdar and A. Bagchi, “On the relation between filter maps and correction factors in likelihood ratios,” *IEEE Trans. Inf. Theory*, vol. 41, no. 3, pp. 833–836, 1995.
- [2] T. Duncan, “On the calculation of mutual information,” *SIAM J. Appl. Math.*, vol. 19, pp. 215–220, 1970.
- [3] D. Guo, S. Shamai, and S. Verdú, “Mutual information and minimum mean-square error in gaussian channels,” *IEEE Trans. Inf. Theory*, vol. 51, no. 4, pp. 1261–1282, 2005.
- [4] P. E. Protter, *Stochastic Integration and Differential Equations*, 2nd ed., ser. Applications of Mathematics. Berlin: Springer-Verlag, 2004, no. 21.
- [5] P. Billingsley, *Convergence of Probability Measures*, 2nd ed., ser. Wiley Series in Probability and Statistics. New York: John Wiley & Sons, Inc., 1999.
- [6] W. Whitt, *Stochastic-Process Limits*, ser. Springer Series in Operations Research. New York: Springer-Verlag, 2002.
- [7] J. Jacob and A. N. Shiryaev, *Limit Theorems for Stochastic Processes*, 2nd ed., ser. A Series of Comprehensive Studies in Mathematics. Berlin: Springer-Verlag, 2003, no. 288.
- [8] H. Kushner, *Approximation and Weak Convergence Methods for Random Processes with Applications to Stochastic Systems Theory*, ser. Series in Signal Processing, Optimization, and Control. MIT Press, 1984.
- [9] S. Ramasubramanian, “A subsidy-surplus model and the Skorohod problem in an orthant,” *Math. Oper. Res.*, vol. 25, no. 3, pp. 509–538, 2000.
- [10] M. I. Reiman, “Open queueing networks in heavy traffic,” *Math. Oper. Res.*, vol. 9, pp. 441–458, 1984.
- [11] R. J. Williams, “Reflecting diffusions and queueing networks,” in *Proceedings of the International Congress of Mathematicians, Vol. III (Berlin, 1998)*, no. Extra Vol. III, 1998, pp. 321–330 (electronic).
- [12] T. Konstantopoulos, G. Last, and S.-J. Lin, “On a class of Lévy stochastic networks,” *Queueing Syst.*, vol. 46, pp. 409–437, 2004.

- [13] H. Kushner, *Heavy Traffic Analysis of Controlled Queueing and Communication Networks*, ser. Applications of Mathematics. New York: Springer-Verlag, 2001, no. 47.
- [14] P. Robert, *Stochastic Networks and Queues*, ser. Applications of Mathematics. Berlin: Springer-Verlag, 2003, no. 52.
- [15] H. Chen and D. D. Yao, *Fundamental of Queueing Networks*, ser. Applications of Mathematics. New York: Springer-Verlag, 2001, no. 46.
- [16] W. Whitt, “The reflection map with discontinuities,” *Math. Oper. Res.*, vol. 26, pp. 447–484, 2001.
- [17] I. Karatzas and S. E. Shreve, *Brownian Motion and Stochastic Calculus*, 2nd ed., ser. Graduate Texts in Mathematics. New York: Springer-Verlag, 1991, no. 113.
- [18] S. E. Shreve, *Stochastic Calculus for Finance I: The Binomial Asset Pricing Model*. New York: Springer-Verlag, 2005.
- [19] ———, *Stochastic Calculus for Finance II: Continuous Time Models*. New York: Springer-Verlag, 2005.
- [20] B. Oksendal and A. Sulem, *Applied Stochastic Control of Jump Diffusions*, 2nd ed. Berlin: Springer-Verlag, 2007.
- [21] M. Zakai, “On mutual information, likelihood ratios, and estimation error for the additive gaussian channel,” *IEEE Trans. Inf. Theory*, vol. 51, no. 9, pp. 3017–3024, 2005.
- [22] D. Nualart, *The Malliavin Calculus and Related Topics*, ser. Probability and its Applications. New York: Springer-Verlag, 1995.
- [23] I. Shigekawa, *Stochastic Analysis*, ser. Iwanami Series in Modern Mathematics. American Mathematical Society, 2004, no. 224.
- [24] T. Kadota, M. Zakai, and J. Ziv, “Mutual information of the white gaussian channel with and without feedback,” *IEEE Trans. Inf. Theory*, vol. IT-17, no. 4, pp. 368–371, 1971.
- [25] R. Liptser and A. Shiryaev, *Statistics of Random Processes I*, ser. Applications of Mathematics. New York: Springer-Verlag, 1977, no. 6.
- [26] C. E. Shannon, “Communication in the presence of noise,” in *Proc. IRE*, no. Vol. 37(1), Jan. 1949, pp. 10–21.

- [27] N. Wiener, *Extrapolation, Interpolation, and Smoothing of Stationary Time Series, with Engineering Applications*. New York: John Wiley & Sons, Inc., 1942.
- [28] M. C. Yovits and J. L. Jackson, “Linear filter optimization with game theory considerations,” in *Proc. IRE*, no. Vol. 43, 1955, pp. 376–376.
- [29] B. Oksendal, *Stochastic Differential Equations*, 5th ed. Berlin: Springer-Verlag, 1998.
- [30] F. C. Klebaner, *Introduction to Stochastic Calculus with Applications*. London: Imperial College Press, 1998.
- [31] D. Revuz and M. Yor, *Continuous Martingales and Brownian Motion*, 3rd ed., ser. A Series of Comprehensive Studies in Mathematics. Berlin: Springer-Verlag, 1999, no. 293.
- [32] B. Fristedt and L. Gray, *A Modern Approach to Probability Theory*, ser. Probability and its Applications. Boston: Birkhäuser, 1997.
- [33] O. Kallenberg, *Foundations of Modern Probability*, 2nd ed., ser. Probability and its Applications. New York: Springer-Verlag, 2002.
- [34] A. Kolmogorov, “On the Shannon theory of information transmission in the case of continuous signals,” *IEEE Trans. Inf. Theory*, vol. PGIT-2, no. 3, pp. 102–108, 1956.
- [35] M. Pinsker, *Information and Information Stability of Random Variables and Processes*. San Francisco: Holden-Day, 1964.
- [36] D. Williams, *Probability with Martingales*. Cambridge: Cambridge University Press, 1991.
- [37] H. L. Royden, *Real Analysis*, 3rd ed. New Jersey: Prentice Hall, 1991.
- [38] I. Gihman and A. Skorohod, *Stochastic Differential Equations*. Berlin: Springer-Verlag, 1972.
- [39] G. F. Simmons, *Differential Equations With Applications and Historical Notes*, 2nd ed. New York: McGraw-Hill, 1991.
- [40] R. F. Bass, *Diffusions and Elliptic Operators*, ser. Probability and its Applications. New York: Springer-Verlag, 1998.